Health Big Data in the Commercial Context

Consumers are increasingly using mobile phone apps and wearable devices to generate and share data on health and wellness. They are using personal health record tools to access and copy health records and move them to third party platforms. They are sharing health information on social networking sites. They leave digital health footprints when they conduct online searches for health information. The health data created, accessed and shared by consumers using these and many other tools can range from detailed clinical information, such as downloads from an implantable device and details about medication regimens, to data about weight, caloric intake, and exercise logged with a smart phone app.

These developments offer a wealth of opportunities for health care and personal wellness. However, privacy questions arise due to the volume and sensitivity of health data generated by consumer-focused apps, devices and platforms, including the potential analytics uses that can be made of such data.

Many of the privacy issues that face traditional health care entities in the big data era also apply to app developers, wearable device manufacturers, and other entities not part of the traditional health care ecosystem. These include questions of data minimization, retention and secondary use. Notice and consent pose challenges, especially given the limits of presenting notices on mobile device screens and the fact that consumer devices may be bought and used without consultation with a health care professional. Security is a critical issue as well.

However, the privacy and security provisions of the Heath Insurance Portability and Accountability Act (HIPAA) do not apply to most app developers, device manufacturers or others in the consumer health space. This has benefits to innovation, as innovators would otherwise have to struggle with the complicated HIPAA rules. However, the current vacuum also leaves innovators without clear guidance on how to appropriately and effectively protect consumers' health data. Given the promise of health apps, consumer devices and consumer-facing services, and given the sensitivity of the data that they collect and share, .it is important to provide such guidance.

To explore the privacy implications of health big data, and to develop concrete proposals for how to resolve privacy issues and at the same time reap the benefits of big data techniques, CDT has undertaken a series of consultations with stakeholders and experts. We examined three scenarios: (1) clinical and administrative data generated by health care providers and payers; (2) health data contributed by consumers using the Internet and other consumer-facing technologies; and (3) health data collected by federal, state, and local governments.

In this paper, we focus on the second of these scenarios: health data collected by non-HIPAA-covered entities through consumer-facing technologies.[1] This includes mobile

[1] HIPAA-covered entities may also collect data through mobile apps, wearables and other interfaces, but that data will generally be covered by the HIPAA framework.

apps, wearable devices, personal health record platforms, social networks, and any other consumer-facing entities outside of the HIPAA framework that collect or share health data relating to individuals. We refer to these as consumer-facing entities, and we refer to their products and services as consumer products. We look both at big data uses by those entities and at their disclosures of data to third parties for research and other analytic purposes.

As the source of privacy guidelines, we look to the framework provided by the Fair Information Practice Principles (FIPPs) and explore how it could be applied in an age of big data to patient-generated data.[2] The FIPPs have influenced to varying degrees most modern data privacy regimes. While some have questioned the continued validity of the FIPPs in the current era of mass data collection and analysis, we consider here how the flexibility and rigor of the FIPPs provide an organizing framework for responsible data governance, promoting innovation, efficiency, and knowledge production while also protecting privacy. Using the FIPPs, rather than proposing an entirely new framework for big data, which could be years in the making at best, would seem the best approach in promoting responsible big data practices. Applying the FIPPs could also help synchronize practices between the traditional health sector and emerging consumer products.

An overarching theme of our analysis is that consumer-facing entities collecting health data about individuals should consider privacy and security when creating their products. Privacy and security protective measures, based on the FIPPs, should be incorporated into the product at early design stages. We detail the steps that developers should take to operationalize each FIPP below.

Openness / Transparency

Openness and transparency should be guiding principles for consumer-facing entities. Fundamentally, it should be clear to a consumer using a health app or wearable device when data is being collected, what types of data are being collected, what that data is used for, what partners it is shared with (and how they use it), how long the data is retained, and what security measures are in place to protect it. Transparency about data practices is essential not just as a fundamental element of privacy but also to engender consumer trust, which in turn is critical to the adoption of these services. Without trust, consumers will resist using apps or devices and the industry as a whole will suffer.

Disclosure about data practices can be done in different ways. At one end of the spectrum, transparency can be provided in a standalone legal notice that provides complete information about information practices. At the other end of the spectrum, practices can be messaged contextually to a user in a way that the user is likely to notice and understand. Both approaches play important roles.

[2] The FIPPs are globally recognized as the foundation for information privacy. There is, however, no definitive version of the FIPPs. We use an articulation of the FIPPs drawn from the Markle Connecting for Health Common Framework, available at http://www.markle.org/health/markle-common-framework and the White House's 2012 Consumer Bill of Rights, available at http://www.whitehouse.gov/sites/default/files/privacy-final.pdf.

Today, unfortunately, standalone privacy policies tend to be inscrutable, risk-averse documents, written by lawyers, in which the primary goal is to be as broad and general as possible.[3] Few consumers read them. However, even these densely written privacy policies do play an important role. The process of drafting a notice can force a developer to fully inventory its data practices. Further, if written clearly and with specificity, they could provide actionable information to those consumers who are particularly interested. Also, they provide a basis for internal and external accountability.[4] Providing technical details about data flows and security practices could also enable commercial entities to compete based on their commitment to privacy and security.

In any case, a notice that is somewhere accessible to users should spell out information about disclosures to third parties and secondary uses (such as analytics or advertising). The entity that directly interfaces with consumers should describe the practices of third parties to which data is disclosed, and those third parties should also have detailed privacy policies posted on their public websites.

Due to the intrinsic sensitivity of health information, commercial vendors have an obligation to clearly disclose data collection practices at a time and in a manner that is likely to be seen and acted upon by the user. Rather than serving only as the basis for user consent, the notices should include concrete, digestible information about what entities actually do with user data. The FTC has made it clear that, even where a registration process obtains express user consent, that consent will be invalid and the data collection illegal if the a reasonable consumer would not be likely to understand the scope of the data practices being conducted.[5]

For mobile apps and wearable devices, notice and choice presents particular problems. Apps may be used on smartphones with relatively constrained display space, and wearables may have even less room. Nevertheless, apps can innovate to adopt very

---

[3] Federal Trade Commission, *What's the Deal? An FTC Study on Mobile Shopping Apps* (August 2014), http://www.ftc.gov/system/files/documents/reports/whats-deal-federal-trade-commission-study-mobile-shopping-apps-august-2014/140801mobileshoppingapps.pdf; Privacy Rights Clearinghouse, *Privacy Rights Clearinghouse Releases Study: Mobile Health and Fitness Apps: What Are the Privacy Risks?* (July 2013), https://www.privacyrights.org/mobile-medical-apps-privacy-alert.

I. [4] The Federal Trade Commission uses its general authority over unfair or deceptive trade practices to take enforcement action against companies that collect or process health data. See, Press Release Federal Trade Commission, "Medical Billing Provider and its Former CEO Settle FTC Charges That They Misled Consumers About Collection of Personal Health Data: Respondents Failed to Inform Consumers They Would Seek Detailed Info From Pharmacies, Insurance Companies and Laboratories" (Dec. 3, 2014) http://www.ftc.gov/news-events/press-releases/2014/12/medical-billing-provider-its-former-ceo-settle-ftc-charges-they ("PaymentsMD case").

[5] *Id.*

simple, very useable interfaces, providing users with clear icons and simple but effective choices. In the case of wearables, the process of connecting the device to the Internet is often the key opportunity to present the user with data disclosure choices.

Notices provided to users as part of the initial activation or registration process might be supplemented by an online document telling a fuller story of data flows. This could help address the concern that a device or app may not be able to fully capture the nuance behind certain practices. By using a short-form notice on the device, and pointing a user to a fuller explanation online if desired, developers can avoid some of the readability problems on devices. Just-in-time notices may also effectively provide information in more digestible and context-specific increments.

How detailed these short-form notices need to be will depend on the context of the relationship with the consumer. Some data collection might be completely obvious and not require a detailed explanation.[6] It should be obvious, for example, that a glucose monitor collects glucose levels; it might not be obvious whom that data is shared with. Companies have an obligation to explain any collection that may not be contextually obvious.

Overall, transparency practices should be guided by the principle that the consumer should not be surprised. The more unexpected or potentially objectionable a data collection or usage would be, the greater is the obligation to explain it to consumers.

Increasingly, it will be possible for companies with *no relationship* with a consumer to collect health information in public spaces; hypothetically, a sensor could be set up on a street corner to monitor the heart rates of passers-by in order to conduct a study of the general population, or potentially to target hypertension ads to relevant consumers. Due to the sensitivity of health information, we believe that it would not be appropriate to collect health information that could be tied back to an individual or a device used by an individual without having a relationship with that individual. In other words, there should be no clandestine collection of health data. Likewise, given the sensitivity of health data, we do not believe that any individual's experience (such as the advertising he receives) should be altered due to observed health information that was not deliberately provided.

The transparency principle also includes the obligation to communicate updates that change collection, use, retention, or security practices.

There are some resources available to assist developers and device manufacturers in formulating their openness practices. CDT and the Future of Privacy Forum released a best practices guide for mobile app developers that highlights the need for effective

---

[6] Federal Trade Commission, *Protecting Consumer Privacy in an Era of Rapid Change: Recommendations for Businesses and Policymakers* (March 2012), http://www.ftc.gov/sites/default/files/documents/reports/federal-trade-commission-report-protecting-consumer-privacy-era-rapid-change-recommendations/120326privacyreport.pdf; The White House, *Consumer Data Privacy in a Networked World: A Framework for Protecting Privacy and Promoting Innovation in the Global Digital Economy* (Feb. 2012), http://www.whitehouse.gov/sites/default/files/privacy-final.pdf.

notice and transparency to users.[7] In 2013, a multi-stakeholder convening organized by the National Telecommunications and Information Administration released a set of principles focused on mobile app transparency.[8] Those principles were later included in an open source privacy policy released by Lookout, a privacy and security startup.[9] Relying on these resources can help developers and manufacturers identify what they need to communicate to users and how.

Commercial privacy policies must be improved to introduce greater accountability for actual practices and these policies should contain detailed information about what data is collected, for what purposes, with whom the data is shared, and how long that data is retained. Companies should also communicate to consumers in simple, clear terms at the time that an app is installed or a device is activated what health information is being collected about the user, and why. How detailed that notice should be will depend on the context of the relationship with the user. In no event should the reasonable user be surprised by the data collection and use.

Purpose Specification and Use Limitations / Respect for Context

Traditionally, the FIPPs were interpreted as requiring entities to specify in advance all the purposes for which data was being collected and to limit future use to those specified purposes unless new consent was obtained. However, some of the most promising applications of health big data are in the field of research; a strict reauthorization requirement could limit future beneficial uses, especially since big data analytics are characterized by their potential to yield unexpected insights.[10]

While developers and device manufacturers should generally try to inform users about the potential for secondary access and use for research purposes, in order to realize the benefits of big data it will not be practical (either in advance or post-collection) to make customers aware of each and every use of their data for research purposes.

The 2012 White House report on consumer privacy uses “respect for context” as a way to describe the essence of the use limitation principle. Respect for context means that consumers have the right to expect that service providers will collect, use, and disclose personal data only in ways that are consistent with the context in which that data was provided. In the context of health big data, commercial entities can use the initial context

[7] Future of Privacy Forum and Center for Democracy & Technology, *Best Practices for Mobile Application Developers*, available at https://www.cdt.org/files/pdfs/Best-Practices-Mobile-App-Developers.pdf.

[8] National Telecommunications and Information Administration, *Short Form Notice Code of Conduct to Promote Transparency in Mobile App Practices* (July 2013), http://www.ntia.doc.gov/files/ntia/publications/july_25_code_draft.pdf.

[9] Ric Velez, Lookout Open Sourced Its “Private Parts,” You Should, Too (Mar. 12, 2014), https://blog.lookout.com/blog/2014/03/12/open-source-privacy-policy/.

[10] As Ira Rubinstein notes, big data “is aimed precisely at … unanticipated secondary uses.” *Ira Rubinstein, Big Data: The End of Privacy or a New Beginning?* (Oct. 2012), NYU School of Law, Public Law Research Paper No. 12-56 (unpublished working paper, available at http://ssrn.com/abstract=2157659).

of data collection as a guide in circumscribing future uses of that data, while still allowing for innovative analytic practices – as long as those uses are related to the initial context.

In the commercial app and device context, the "respect for context" principle is an especially helpful way to contemplate potential big data uses that will not surprise users but will allow for future research and serendipitous discovery. Internal operational research – such as security improvements, stability refinements, and future product development – should reasonably qualify as secondary uses of data that fall within appropriate contextual use. Users of health apps and devices should expect that developers and manufacturers will use the data collected in order to improve the functionality of their services and to develop potential new features in keeping with the original stated purpose of the app. If the app developer or device manufacturer contemplates functionality that goes further afield of the original purpose of the app (for example, changing a blood sugar tracking app into a general health data sharing service), that would fall outside of the original context and should prompt the developer to seek reauthorization of consent from consumers.

While internal research use may be permissible, developers and manufacturers should still strive to ensure that their users understand that their data may be accessed and used internally for these types of secondary purposes. They should provide information about operational secondary uses in their formal privacy notices and in other detailed online notices. There should also be an effort to retrospectively message back to users representative examples of knowledge gleaned from research with users' data.

On the other hand, the transfer of personal health information in identifiable form to third parties for secondary purposes does not respect context, but rather would surprise most patients. Identifiable data shouldn't be shared outside an organization for either research or operational use unless an agreement is in place that limits uses to the primary purpose and applies the same (or higher) level of protections to the practices of the third party. (As we discuss below in the section on Data Minimization, it might be acceptable to transfer health information in identified form to be merged with other data for research purposes if there is an immediate commitment to de-identify the merged data set.)

Even when health data is not transferred to a third-party, secondary uses may raise concerns; in particular, manufacturers and developers should be wary of using health data on behalf of external entities, especially for marketing purposes. Providers should not send marketing offers on behalf of third parties — targeted by information contained in a user's data record — unless the patient affirmatively opts to receive these offers.

We also suggest three other criteria in determining whether a secondary use may not require new permission from the user: (1) where secondary use of health information should only occur where no extra data collection, retention, and transfer occur (apart from transfer to dedicated service providers with no independent right to use the data, (2) where the secondary use of the data is not used to materially alter the end user experience, and (3) where no human observes or processes personalized information but instead the data is processed by a machine, and the output is aggregate and not personally identifiable.

We argue above that internal operational uses for product improvement purposes are generally consistent with user expectations, and providers should not have to call outside a privacy policy this purpose and get dedicated permission. In this context, service providers may engage in A/B testing (presenting different information or formats to different users to test reactions), and in most cases such practices should not require express user consent. However, it has been disclosed that some online services conducted tests that deliberately delivered to some customers a negative experience.[11] For A/B testing or similar research using health information, at the very least companies should follow the doctrine of "first, do no harm." That is, companies should not deliberately give users an experience they think will be negative. Rather, A/B tests should be limited to alternatives that the provider believes in good faith are both equally good for the user.

In determining whether research or analytics is consistent with the context in which data was initially collected or requires additional consent, companies should consider multiple factors, including (1) whether the research is done internally (or by third parties acting under the same limits that apply to the consumer-facing entity), (2) whether it materially changes the experience of the user, (3) whether it involves human processing; and (4) whether it only yields de-identified, aggregate results. Companies could increase trust in their research uses by retroactively telling users about test results.

Focused Collection / Collection Limitation

Especially considering the sensitivity of health information, limiting collection in the health app and device context remains of prime importance. While some proponents of big data argue that the full benefits of big data cannot be realized without unbridled collection, we believe that focused collection builds consumer trust, without which societally beneficial commercial applications that access and use health information are unlikely to be adopted.

When using an app or device for health purposes, a consumer will of course assume that health data will be collected. But it is important that developers and manufacturers collect *only the data required* for the fundamental purpose of the app. When adopting a health app or device, consumers hope that they can improve their health and well-being, and they provide sensitive information in order to do so. To ignore that expectation by over-collecting data that might one day be relevant would violate the trust proposition offered by developers and manufacturers. Of course, life logging applications and personal health records should be configurable to store whatever a patient wants. But companies should set reasonable defaults based upon the nature of the product and perceived consumer expectations, while providing controls that allow individuals to log and share only the data elements they want.

[11] See Brian Fung, *OkCupid reveals it's been lying to some of its users. Just to see what'll happen*, Washington Post, July 28, 2014, http://www.washingtonpost.com/blogs/the-switch/wp/2014/07/28/okcupid-reveals-its-been-lying-to-some-of-its-users-just-to-see-whatll-happen/.

To that end, commercial entities offering health products to consumers should not collect information about users in ways that would surprise the consumer or violate the trust of the relationship. A smartphone app designed to help users monitor their caloric intake, for example, should not collect location information from the phone, even though that could possibly help track what restaurants the user visited. Rather than collecting every piece of potentially valuable information without restraint, developers and manufacturers should consider the context and the consumer's reasonable expectations to determine what information should be collected.

Potentially, health app developers and device manufacturers could obtain data from other sources (such as from data brokers, public records, or business partners) and append that to the data that they collect from the app or device. While this may tempting in order to facilitate further big data uses, supplementing sensitive datasets with outside information would probably run counter to ordinary consumer expectations. Should developers and manufacturers feel the need to supplement their records with additional information, they should approach users directly and seek consent. Commercial vendors should adhere to a practice based on HIPAA's "Minimum Necessary Requirement," where only the health information that is necessary to perform a particular function is collected.

Data Integrity and Quality / Data Access and Accuracy

The importance of data integrity and accuracy in the health context is obvious. As health care becomes more data-dependent, inaccurate, outdated data could have major adverse repercussions. Less apparent may be the accuracy and reliability issues associated with data analytics. A wide range of entities use algorithms to support important processes, including in the health context. The widespread use of algorithms – which are often obscure to ordinary consumers — highlights the increased need not only for accuracy of data but also for the reliability of the analytic processes applied to that data.

Closely associated with the principle of data accuracy is the principle of access: individuals should be able to access and copy data about themselves, both to check its accuracy and to use it for their own benefit. In the HIPAA context, progress is finally being made, through the Blue Button initiative, in providing patients with meaningful access to their electronic health records. In the commercial context, all consumer-facing apps should be built on the premise of user access and portability. Presently, users of apps and devices do not always have an affirmative opportunity to view their records. If developers and manufacturers gave consumers access to their data and periodically prompted them to do so, with convenient tools and interfaces, integrity and trust would benefit. "If organizations provide individuals with access to their data in usable format, creative powers will be unleashed to provide users with applications and features building on their data for new innovative uses."[12] The Blue Button framework may make

[12] Omer Tene and Jules Polonestsky, Big Data for All: Privacy and User Control in the Age of Analytics, 11 Nw. J. Tech. & Intell. Prop. 239 (2013), available at http://scholarlycommons.law.northwestern.edu/njtip/vol11/iss5/1.

it easier for developers and manufacturers to allow patients to download data into tools of the patient's choosing, and to directly transmit this data to other entities.[13] It may be reasonable to charge a fee for this service, though the fee should not be prohibitive, and should only cover the reasonable costs associated with maintaining the access feature.

Likewise, consumers should have insight into the algorithms that are applied to their data, in order to avoid possible negative outcomes, such as discriminatory treatment or the drawing of incorrect assumptions. (In the commercial context, it may not be appropriate to make available to consumers all details about the algorithms involved as they may be proprietary trade secrets. In the context of big data, the principle of data integrity or accuracy should encompass the reliability or accuracy of outcomes. Researchers should be cognizant of the risks of algorithm-based analysis. There is in fact a growing body of knowledge around the risks of big data analytics that should be imported into the health data space.[14] Developers and manufacturers of consumer apps and devices should develop auditing procedures to assess the reliability of results generated by their analytics.

The accuracy principle should apply not only to data but also to the analytic processes applied to data and the outcomes generated by such analytics. We believe that companies developing consumer apps and devices should ensure that consumers can easily access their data and copy it in portable formats.

Data Minimization

Generally, the data minimization principle requires entities to collect no more data than is necessary for the purpose at hand and to delete data when it is no longer needed for that purpose (or a contextually related one). However, data minimization has been positioned as antithetical to the goals of big data. If entities are required to minimize collection and delete data if no longer necessary, the thinking goes, they will be unable to realize the potential upsides of big data analytics — especially future research not contemplated at the time the data was collected. However, this viewpoint ignores the risks that retained data sets pose: the potential for data breach or misuse, as well as the potential chilling effect on consumers unsure about what will happen to information they give to a service provider.[15]

Individuals who use life-logging applications or personal health records must be in control of the data they are creating. Certainly, individuals should be entitled to share as much data as they want if sufficiently informed. However, companies offering consumer-

[13] Deven McGraw, Helen R. Pfister, Susan R. Ingargiola, and Robert D. Belfort, Lessons from Project HealthDesign: Strategies for Safeguarding Patient-Generated Health Information Created or Shared through Mobile Devices, 26 J. Healthcare Info. Mgmt. 25 (2012), available at https://cdt.org/files/pdfs/JHIM-Lessons-from-Project-HealthDesign.pdf.

[14] Solon Baracos and Andrew Selbst, *Big Data's Disparate Impact*, February 13, 2015, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2477899.

[15] Justin Brookman and G.S. Hans, *Why Collection Matters: Surveillance as a De Facto Privacy Harm*, http://www.futureofprivacy.org/wp-content/uploads/Brookman-Why-Collection-Matters.pdf.

facing services should offer users the means to delete their data whenever they want, both complete records as well as individual data elements (quite different considerations apply to data held by providers and payers). Users should be able to remove health information from public (or private) view and also from a company's records. While a company may not be able to locate all copies of the data, it should undertake a good faith effort to delete where possible, and should commit not try to recreate the dataset in the future.[16]

For services that are not used by a consumer to store health information but that merely observe transactions from which health information could be inferred (such as buying a medical device through an e-commerce site, or searching for a medical condition on a medical information site), requiring a right of deletion is a closer question. Still, to the extent that a service can authenticate a particular user, we believe that it is a best practice to let that user delete derived health information, unless necessary for an essential business purpose. Better yet would be to design systems so that they do not collect data that could be linked to an individual. Sites for medical information, for example, should carefully design their data collection polices to limit third-party sharing — as well as its own use of cookies and other identifiers — to what is operationally necessary.

De-identification may allow companies to retain datasets for research and analytics beyond the time when the minimization principle would otherwise call for deletion. That is, data may be used and shared for research purposes if it has been meaningfully de-identified such that the information could not likely be traced back to a specific user. The de-identification test promulgated by the Federal Trade Commission for data not regulated by HIPAA provides a flexible framework to consider for health app developers and device manufacturers. That test states that data is not "reasonably linked" to an individual or a device if (1) the party takes reasonable measures to de-identify the data, (2) commits to not re-identify data, and (3) prohibits downstream recipients from re-identifying the data.[17]

De-identification cannot be presumed to eliminate all potential risk of re-identification of data relating to individuals. [18] Because de-identification does not eliminate risk of re-identification, protections are still needed for the residual re-identification and other privacy risks that remain in the data.[19] However, requiring reasonable assurance that the

---

[16] Center for Democracy & Technology, Comments to the National Telecommunications and Information Administration on "Big Data and Consumer Privacy in the Internet Economy," August 5, 2014, https://cdt.org/insight/comments-to-ntia-on-big-data-and-consumer-privacy/.

[17] Federal Trade Commission, *Protecting Consumer Privacy in an Era of Rapid Change: Recommendations for Businesses and Policymakers*, March 2012, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2477899.

[18] Ann Cavoukian and Daniel Castro, Big Data and Innovation, Setting the Record Straight: De-identification Does Work (June 16, 2014), available at http://www2.itif.org/2014-big-data-deidentification.pdf; Arvind Narayanan and Edward W. Felten, No Silver Bullet: De-identification Still Doesn't Work (July 9, 2014), available at http://randomwalker.info/publications/no-silver-bullet-de-identification.pdf.

[19] Deven McGraw, *Building Public Trust in De-Identified Health Data*, 20 J. AM. MED. INFO. ASS'N 704 (2012), *available at* http://jamia.bmj.com/content/early/2012/06/25/amiajnl-2012-000936.full.html.

data could not be re-identified allows societally valuable data sharing to occur while mandating strong technical and policy protections to prevent reattribution.

One way to reduce the risk of harmful re-identification, of course, is to limit the collection of data so that the individual data record does not contain extraneous identifying information that could be re-identified. By doing so, the potential privacy risks are mitigated.

Another challenging question is whether identifiable data sets can be ephemerally merged and then immediately de-identified. The initial merger would violate the general prohibition on transferring identifiable health information absent consent; however, the commitment to immediately de-identify would mitigate any risks associated with such a transfer. While allowing for data merger coupled with de-identification does increase the potential for accidental data exposure, we believe the benefits that could accrue from expanded databases for research may outweigh these privacy risks. We are inclined to support an exception on the prohibition on transferring identifiable information absent consent where there is a commitment to immediately de-identify data, a policy consistent with the FTC test described above.

Companies should not collect health information that is not necessary for a purpose clearly articulated within a privacy policy at the time of collection, and they should delete identifiable copies of such data when it is no longer necessary for that purpose. Individuals should have the ability to delete health information that commercial entities store about them, especially data in personal health records created by the consumer or and in life-logging services. However, data that has been reasonably de-identified may be retained and shared for research so long as the company and recipients of the data commit to not re-associate the data with a particular individual or device. We also tentatively support allowing companies to merge identifiable data absent consent so long as there is a commitment to immediately de-identify the merged data, prior to conducting any analysis of the new data set.

Individual Participation / Control

Fundamentally, the decision to share or transmit health information generated by a commercial app or device must reside with the individual. When using a commercial app or device, users should feel that they control their data, rather than merely sending it off into the cloud to be analyzed, modified, and shared with third parties. By promoting user control, developers can also promote trust in their app or device, which is especially crucial in the health context.

Individual control is necessarily connected to the notions of transparency and notice discussed above. Secondary uses such as internal stability research and security development may not require individual control through authorization and re-authorization from users, but developers and manufacturers should institute strong internal audit and oversight procedures to prevent internal misuse.

Some commentators have pointed to the ever-expanding nature of data flows as a rationale for diminishing the role of individual control.[20] We do not necessarily agree. Ultimately, users must have the capacity to regulate the flow of their personal health records and other data collected through apps and devices outside the clinical context.[21] Service providers should set reasonable defaults for information collection and retention, but the ultimate decision about what to collect, retain, and share should lie with the individual. A user may not need to be asked about every potential secondary usage, but for some, such as first-party marketing based on health information, the user should have the ability to opt out.

On the other hand, the transfer of data to third parties for marketing, advertising, and even research should require opt-in consent by individuals, unless those third parties are dedicated contractors providing services essential to operation of the product with no independent right to use that data. This consent should not merely be a blanket permission contained within an opaque terms of service agreement, but should reflect the user's will, either to freely provide that data, or to provide it in exchange for a service. In doing so, developers and manufacturers will encourage patient trust regarding secondary uses, and allow for individual participation for such uses. Without such reassurances from providers that privacy and security considerations are being kept in mind, patients may be wary of using health apps and devices. As noted above, however, the use and disclosure of de-identified data should not require individual consent and control.

We are not convinced that entities with which a user has no direct relationship should be collecting identifiable health information. While we believe that information should typically be collected with consent, there may be exceptions for third-party data collection not for targeting or research purposes, but for security and fraud prevention (such as for the prevention of click fraud in online advertising). At the very least, passive, third-party data collection and use — such as by online marketing companies should be subject to some sort of pervasive, universal controls as determined by the user. In the online environment, stakeholders have attempted to create those tools through a universal "Do Not Track" setting though to date advertising companies have not agreed to honor those settings. The White House recently encouraged the development of these sorts of similar tools in its Big Data review.[22] Widespread adherence to a workable, user-centric Do Not Track standard for online information collection would be one way to enhance user control across the entire Web and would be especially valuable for health-related sites..

---

[20] *See e.g.* Craig Mundie, Privacy Pragmatism, March/April 2014, http://www.foreignaffairs.com/articles/140741/craig-mundie/privacy-pragmatism.

[21] Simone Fischer-Hübner et al., *Online Privacy – Towards Informational Self-Determination on the Internet*, August 28, 2013, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2468556.

[22] Executive Office of the President, *Big Data: Seizing Opportunities, Preserving Values*, May 1, 2014, http://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf.

Personal health information should only be collected and shared with at the direction of the individual. With limited exceptions, passive collection of health data should not be conducted, given the sensitivity of such data and the potential for misuse.

Security

It is now well-established principle that any entity collecting individual data is responsible for protecting the security of that data. Recent high profile breaches and enforcement actions by the FTC reemphasize the need for strong security. A reasonable program must address technical, administrative and personnel measures and must include regular auditing and frequent updating.

Given the intrinsic sensitivity of health information, developers and manufacturers have heightened security obligations beyond the digital ecosystem at large. In order to incentivize strict privacy and security protections, there must be significant consequences for stewards of data when that data is misused or illegitimately accessed. A majority of states have laws requiring data holders to notify users of breaches. These breach notification requirements perform several functions. First, they provide transparency to patients that their information has been illegitimately accessed, empowering them to take precautionary actions to limit potential consequences (such as identity theft). Second, and perhaps more importantly, breach notification requirements impose significant costs on the service provider — both in terms of actual costs in distributing notices as well as lost public goodwill — which provides a strong incentive to safeguard the data in the first place.

Encryption is an important element in protecting health data. While entities in the consumer app and device context are likely not covered by the HIPAA Security Rule, its focus on encryption is a helpful starting point for developers and device manufacturers to consider when designing their security programs.[23] Developers and device manufacturers should encrypt data both when it is stored on servers and devices and when it is in transit. Should default encryption not be feasible, security measures such as de-identification, limited transmission of unencrypted data, and collection and retention limitations would help limit risk. Ensuring that systems and processes are up-to-date, as well, is vital in creating an appropriately robust security program.

FTC interpretation of its Section 5 authority already requires reasonable security for all personal information, including health information. However, it is not clear that the need for strong data security has sufficiently internalized by corporate decision-makers. Companies should have robust data security plans in order to force companies to consider security threats and to put into place reasonable measures proportionate to the security risk and the sensitivity of the health information.

Accountability, Oversight, Remedies

[23] Deven McGraw et al., *Lessons from Project HealthDesign: Strategies for Safeguarding Patient-Generated Health Information Created or Shared through Mobile Devices*, Journal of Healthcare Information Management, Summer 2012, https://cdt.org/files/pdfs/JHIM-Lessons-from-Project-HealthDesign.pdf.

As the size and diversity of datasets grow, and as analytic techniques make it easier to re-identify data and draw inferences from seemingly innocuous data, internal and external accountability must play a larger role in protecting health data. Whenever data is illegitimately used or transferred — or reasonable security protections are not put into place — regulators and patients must be able to obtain compensation and punitive remedies from bad actors.

Currently, the Federal Trade Commission uses its basic authority to bring enforcement actions against companies that fail to adequately protect user data. The FTC has targeted companies that over-collect information which use data in ways or that are inconsistent with stated privacy policies or terms of service, and that fail to take reasonable steps to safeguard data. Undoubtedly, it will bring enforcement actions on similar grounds in the health context. (The FTC could be even more effective if it had wider authority to impose penalties for privacy violations.) State attorneys general also have investigative and enforcement powers.

Internally, in addition to having a robust security program, companies that handle health information should also have privacy processes in place to ensure that products are developed with privacy and the primacy of the user in mind. We have previously advocated for strong access controls that limit unauthorized access to personal health data;[24] this is a vital step in promoting privacy. The appointment of a chief privacy officer, even for start-ups, can provide a focal point for internal privacy policy development and compliance.

Companies collecting health data should adhere to the Fair Information Practice Principles described above and they should have processes in place to assess their compliance with their own rules as well as any applicable laws. They should also have formalized internal processes to ensure that privacy-conscious decisions are made through a product's lifecycle and that analytical decisions and determinations made using consumer-generated health data are accurate and fair.

Overall, it's critical for commercial entities that collect, use and share personal health information to imbed strong privacy and security standards into their products and services in order to maintain customer trust, improve adoption and retention rates, and provide the myriad potential health benefits that can come from empowering consumers with their health data.