

Big Data, Big Questions: Implications for Competition and Consumers

September 21, 2021

Response to Questions for the Record

Charlotte Slaiman

Competition Policy Director

Public Knowledge

Washington, D.C.

Responses to Senator Grassley

1. How important is the amount of data that a company has to their ability to effectively monetize that information?

One of the important features of data as an economic good is that it exhibits increasing returns to both scope and scale. That means that more data is exponentially more valuable, particularly when it comes from different sources. This makes it incredibly difficult for a startup company with only one source of data to properly compete against Big Tech's massive data collection operation.

2. How difficult can it be for a startup or small business to collect enough data to be able to compete with companies that have large amounts of data?

It can be very difficult. Consumers might have less trust in a name or brand they do not recognize. An already established Big Tech titan can use their other sources of data on you to "fill in the gaps" in a way that a smaller business with just one or two products cannot, unless they purchase that data from an external source. This allows for a greater degree of targeting and other data exploitation by Big Tech firms.

3. Some commentators argue that the amount of data currently possessed by large incumbent companies forecloses the ability of new entrants to compete. But, new data is being created every day and what data is important in the future may not be what is being collected today. If so, why isn't there an opportunity for additional companies to enter the market?

There is still not a fair opportunity for additional companies to enter the market because of the powerful market position of the largest platforms, together with the lack of fair competition rules such as interoperability and non-discrimination requirements. This allows the largest platforms to continue their control of new data points and makes it much harder for competitors to enter.

I addressed this question further in my written testimony on page 7:

“The dominant platforms sometimes argue that data ages rapidly, so a new competitor could quickly amass the data needed to compete. This is theoretically true, but it is not just old data that today’s dominant digital platforms control. Users are still today locked in to these platforms through their gatekeeper power, so the platforms continue to have access to ongoing data streams. These data streams can be used to continuously update algorithms to stay on top. They can also keep platforms updated about the users so as not to rely on just static or past data about them. This allows platforms to not just ‘know’ their users, but see how their users change over time—often in response to the algorithms that platforms created using older user data. This cycle of collection, use, and iteration gives platforms significant power over their users.”

4. There is debate over whether Big Data should be regulated through the lens of consumer protection and privacy or whether antitrust laws should be used to address competition concerns with the collection of data. Do you have an opinion about the best approach to address this issue or should we be looking at a combination of different approaches?

I certainly believe that a combination of approaches is superior. Technology platforms present a myriad of issues for our society, and our solutions need to account for the interactions at play. Antitrust alone is not enough. There is a needed role for privacy, consumer protection, and civil rights. A privacy regime focused on data minimization and a more restrictive environment for the collection, spread, and use of consumer data would certainly help control for some of the advantages dominant platforms receive from their unchecked and pervasive collection of consumer data. Public Knowledge is supportive, for example, of bipartisan comprehensive privacy legislation that Congress is considering that would limit data collection practices of online platforms, data brokers, and others.

Consumers should not have to choose between competitive markets and their privacy. Congress must act to safeguard user privacy *and* to promote fair competition on and against the largest firms.

Responses to Senator Tillis

1. The term “data” can have multiple meanings, which can sometimes generate confusion in policy discussions. How would you define “data” and “big data”, as used in your written testimony?

Your question gets at the complexity of the “data” definition in policy discussions. This can be a difficult question to answer and is perhaps why recent legislation (S. 2992, *The American Innovation and Choice Online Act*) calls for a Federal Trade Commission rulemaking to define “data.” I would define “data” as anything a platform collects and uses to either monetize or improve its services. However, the term certainly can have multiple meanings, so it’s important

that we continue to define what we mean by “data” in different contexts. “Big data” refers to data sets large enough to be useful for training machine learning algorithms and similar purposes. Big Tech platforms have a built-in advantage in collecting big data sets.

a. How would you define consumer and user data, specifically what would be included and excluded from these definitions? For example, the section of data portability refers to “your data” – what would this include?

When it comes to protecting user privacy, consumer and user data can be defined as information that identifies or is linked or reasonably linkable, alone or in combination with other information, to an individual or a device that identifies or is linked or reasonably linkable to an individual and may include derived data and unique identifiers. This could be things like your IP address, your mobile phone’s device identifier, location data, or anything collected about your personal behavior online or when combined with information collected online is reasonably likely to identify an individual. In the context of data portability, the definition of “your data” would be different for different product categories based on what is needed to promote a competitive marketplace. Ongoing communication across social media networks requires interoperability of different categories of data than porting a customer over to a rival phone carrier, for example. Of course, since portability or interoperability should only be happening at the user’s request, the user would also decide which data categories they actually want to transfer.

b. Would this include user uploaded videos, images, and text?

Yes, in general I would expect effective data portability requirements to include these items. This means companies would need to offer this option to consumers, and consumers could choose which of their videos, images, or text communications—if any—to transfer.

c. Would such content be considered part of the “user” data, even if it includes content that originates from other sources?

Yes. A user should be able to download material they have uploaded, even when that content originates from other sources. Uploaded material (for example, in the form of memes, reaction images, or even emoji) is as much a part of how many people communicate online as their own words. Failing to allow this material to be exported would hamper the goals of competition and interoperability.

d. Does it include data in which intellectual property rights, including copyright, trade secret, trademark, or design rights, may subsist?

Not necessarily. I would think that the important part of most user’s data is their relationship to a copyrighted work, not necessarily the work itself. For example, my user data might include positive interactions with the movie “Mad Max: Fury Road”—posts I’ve liked about it, me posting about it, etc. However, that data would not implicate any sort of IP rights in the actual movie. In some cases, users may upload content (e.g., memes or snippets of text) that are fair

uses of copyrighted material. In any event, data portability requirements do not provide new grounds for the infringement of IP rights.

2. Your testimony advocates for “a digital regulator to comprehensively the policy questions surrounding digital platforms.”

- a. Given the many issues beyond privacy and competition that address and implicate digital policy—including cybersecurity, national security, consumer rights, free speech, and intellectual property concerns—what existing agency would be the best situated, in your view, to carry out this role?**

There is no one agency today that is best suited to properly deal with the cross-cutting issues presented by technology platforms that your question identifies. That is why Public Knowledge advocates for a completely new agency. Both the Federal Trade Commission and the Federal Communications Commission have some tools and expertise in common with what a digital regulator would need—but either one would need additional authority, staffing, and budget to achieve the goals of a digital regulator.

- b. Is it important to you that the regulator should be politically accountable?**

Political accountability is important and can foster trust in an agency’s decisions. Public Knowledge has advocated for a bipartisan commission-like structure (similar to the Federal Trade Commission or the Federal Communications Commission) as the structure that strikes the right balance between political insulation and accountability.

3. Your testimony refers to data portability as an important tool “to neutralize the power that Big Data confers upon dominant digital platforms.”

- a. In your view, is the Data Transfer Project, described in Google’s testimony as a partnership among Google, Apple, Facebook, Microsoft, and Smugmug, an acceptable way to address data portability?**

While projects like the Data Transfer Project are laudable, I believe they are insufficient to solve data portability and interoperability problems. Company commitments without agency enforcement only get you so far. Firms might not make some key data available and can cut off portability to rivals that pose a competitive threat. This is a key component of the ongoing Facebook FTC litigation. Facebook offered access to its network to companies that would increase engagement with the core Facebook product but cut off would-be rivals like Vine. I strongly believe we need federal legislation in this area in the vein of the bipartisan ACCESS Act.