DIGITAL DATA AS AN ESSENTIAL FACILITY: CONTROL

BY CATHERINE TUCKER[1]

1 Catherine Tucker is the Sloan Distinguished Professor of Management Science and Research Associate at the NBER. Please see my disclosure statement at https://mitmgmt-faculty.mit.edu/cetucker/disclosure/.

I. ARE DATA AN ESSENTIAL FACILITY?

The notion of an "Essential facility" is key to antitrust law, because denial of access to key facilities can mean that a potential monopolist will be immune, at least for some time, to most forms of competition. As an economist, I observe that legal scholars debate whether anything can really be an essential facility, given that one can always raise exceptions or caveats to any even potentially watertight case.

Given this scholarly debate, it is useful to introduce terminology used for a parallel notion in strategic management. Here the focus is on a "resource" that gives "sustainable competitive advantage," or that can act as a barrier to entry. The idea of such a "resource" was first articulated by my colleague Birger Wirnerfelt at MIT,[2] and then further refined by Barney (1991).[3] To be a "resource," the asset in question needs to be valuable, non-imitable (or at least difficult to substitute), and rare. More recent theory has introduced the additional nuance that the firm must be able to "control" this rare, valuable, imperfectly imitable and non-substitutable resource.[4]

When we teach this framework, our aim is to give managers a means to consider what resources a firm has, which may allow them to build sustainable competitive advantage in such a way that would allow them to profit. However, a parallel use of this framework can also be to allow an examination of potential barriers to entry, and, at the extreme, whether or not a firm's "resource" or "core competency" has morphed to the extent that it could be considered to be an essential facility.

In earlier work I went through the first four pillars of this strategy framework (i.e. the conditions relating to a given facility of asset being rare, valuable, non-imitable and non-substitutable) in considering whether data could confer such an enduring competitive advantage.[5] In this essay, I will consider the more recent additional pillar that has been added to this strategic framework – that is, the question of whether firms really control data in a way that makes it a source of competitive advantage.

2 Wernerfelt, Birger, "A resource-based view of the firm," Strategic Management Journal 5.2 (1984): 171- 180. The original article has been cited over 33,000 times on Google scholar. It was also popularized by Prahalad, C. K. & G. Hamel, "The core competence of the corporation," Harvard Business Review 68.3 (1990): 79-91.

3 Barney, Jay B., "Firm Resources and Sustained Competitive Advantage," Journal of Management 17.1 (1991): 99-120. This paper has over 70,000 Google Scholar citations.

4 Barney, Jay, Mike Wright & David J. Ketchen Jr., "The resource-based view of the firm: Ten years after 1991," Journal of Management 27.6 (2001): 625-641.

5 Lambrecht, Anja & Catherine E. Tucker. "Can Big Data protect a firm from competition?," Competition Policy International (2015).

II. INTERNAL MEANS OF CONTROL OVER DATA

The type of data that firms can directly control access to differs according to its purposes and between firms. Typically, the key distinction is whether the presentation of data is internal or external. For example, YouTube cannot easily prevent access to data about who has liked a video on YouTube, or the content of comments. However, Uber can choose not to make public data on the geographic granularity of rider pickups. Of course, even in that case it is somewhat possible to recreate the data, and indeed a French PhD student has managed to access partial data through the Uber API for her work on gender discrimination in ride pickups.[6] In general, though data is non-rivalrous, it is possible to exclude access to particular data if the data are not public. Sometimes, the legal treatment of data has focused on the idea of non-rivalry – which is indeed a key component of the definition of a public good – without also acknowledging that much of the time the same digital tools that allow the collection of vast datasets also permit control over who accesses it.

The key point is that firms are often able to control access to a particular dataset. What they are far less able to do is to control the ability of rival firms to create a similar dataset. In particular, they are unable to control the ability of a rival firm to create a dataset which offers similar insights. This latter point is important because ultimately the value of data is not the raw manifestation of the data itself, but the ability of a firm to use this data as an input to insight.

In general, I argue that the ability of a firm to control whether a rival also creates a similar dataset very much depends on the extensiveness of the digital footprint that a consumer has when generating data that gives that particular insight.

Let us contrast a few scenarios which are all taken from the context of digital advertising. I recognize that the power of digital data may well be greatest outside the world of digital advertising (for example in optimizing logistics). I focus on digital advertising simply because that is the focus of much of the current antitrust debate involving data in this context:

1. A firm owns data allowing them to identify both the income and zip code of a particular set of "eyeballs" arriving at a website, and thereby to know what hotel ad to show to that set of eyeballs.[7]

2. .A firm owns data that allows them to predict what smartphone a set of "eyeballs" is most likely to buy, and to feature this smartphone in personalized recommendations.

3. A firm owns data that identifies that a set of "eyeballs" is shopping for last-minute flowers to be delivered on Mother's Day.

As you read this essay, it might be useful for you to rank these examples in terms of likelihood of the firm being able to "control" this data as a source of sustained competitive advantage. I would argue that in each instance, there is a feature of the data which limits control.

A. When Data Reproducibility Limits Control

In the Stigler report, a motivating example for why data might have increasing returns is the idea that one potential source of data about hotel bookings is locational data.[8] The second useful source of information might be presumed income of the individual. The Stigler report argues that the combination of data about zip code and income may be more powerful than the information about each data point individually. However, what is striking about this example is that the two sources of data (zip code and income) are publicly available, and are therefore not hard for any firm to gain access to.

6 "Algorithm Bias on Uber" – Clara Jean, Université Paris Sud, Working Paper (2020).

7 This is a scenario discussed in the Stigler Committee on Digital Platforms Final Report (2019), https://www.publicknowledge.org/wp-content/uploads/2019/09/Stigler-Committee-on-Digital-Platforms-Final-Report.pdf.

8 In the Stigler report it states that the "home zip code" of the user is useful as there may be some "limited ability to predict interest in a hotel based on this zip code being sufficiently far from the hotel." The report formally models the likelihood of a user booking a hotel as a function of their distance from the zip code. I assume this means this implies that I am most likely by their model to book a hotel in Augusta (Western Australia), which is the city I believe furthest from Boston. Interestingly, before reading the Stigler report prediction that I would book a hotel in Augusta, I had not contemplated going to Augusta. I agree it does look like a nice destination, though given that Augusta only has two hotels, I suspect that the prediction task is not that hard.

The fact that data is non-rival and virtually costless to produce has led to a large industry of data brokers such as Acxiom and Experian, who collect and parse data about people's internet activity. This is usually the focus of discussion about data aggregation and reuse. However, this example of location and income information is a useful example of the types of data which are freely available through public datasets. It is reasonably easy to geo-locate any set of eyeballs using a digital device.[9] It is also reasonably easy to use census data to obtain the average income of a zip code.[10]

Given how relatively easy it is to reproduce zip code and income information from public sources and then link them to the geographical location of a user, then this will inherently limit a firm's ability to control such data as a source of sustained competitive advantage.

B. When the Ability to Recreate the Insights of the Data Limits Control

In the second example, it is notable that the potential source of competitive advantage is not data *per se*. Instead, it is the prediction based on data about what smartphone someone is likely to purchase. Indeed, no company knows for certain what smartphone someone will buy.[11] There are many potential ways to predict smartphone purchase likelihood.

One potential course is to use paid search data on the type of phones someone has searched for. Another way might be to use browsing data on a review site to see what kind of reviews the user is browsing. Another way may be to use "look-a-like" data to study what smartphones other similar people have bought in the past.

One issue is that this kind of prediction task is not easy. I have studied this in my own research – (Neumann et al., 2019) – where we show that often predictions of people's age and gender based on browsing data are not much better than random chance.[12] However, that research seems to suggest that it is the skill of the humans in charge of the predictions that seems to matter, in that prediction improves when a human guides the choice of the algorithms.

Therefore, the key questions are, whether there is only one set of data that is appropriate for a prediction task; and also to give appropriate credit for the difficulty of creating insights from the data, rather than ascribing the rarity of the insight to the data itself.

C. When the Short-lived Nature of the Data Limits Control

I chose the example of the Mother's Day flowers, because this is an instance where the consumer is likely to have a small digital footprint. Someone running late for Mother's Day is likely to perhaps only engage with one search engine or one website when searching for flowers. As a result, it is possible that only one firm has access to the insight that someone is wishing to buy flowers for Mother's Day. This enhances a firm's potential control of the data.

However, there is a caveat, which is that in this particular example the value of the data is very short-lived. Once Mother's Day is over, the data loses its value, rendering it not a good source of sustained competitive advantage. Next Mother's Day, the firm will have to compete again to be the digital platform or website which gains the insight. The only chance of the data retaining its value is that the same person may be tardy at buying flowers on Mother's Day the next year, such that the data become useful as an annual prompt.

This example shows that it is hard to generalize about how easy it is to `control' the value of any one bit of data. However, it also shows that it is important when thinking about control of data, to also think about how long-lived any insight from the data might be.

---

9 For example, one can access this information for browsers from https://www.w3schools.com/ html/html5_geolocation.asp or via other platforms such as https://developers.google.com/web/ fundamentals/native-hardware/user-location. There are of course issues with accuracy if people take steps to deliberately mask their location, but these are likely to be shared by all firms.

10 Go to https://catalog.data.gov/dataset/zip-code-data.

11 I am not sure what smartphone I will buy next.

12 Neumann, N., C. Tucker & T. Whitfield (2019). "Frontiers: How Effective Is Third-Party Consumer Profiling? Evidence from Field Studies Nico Neumann, Catherine E. Tucker, and Timothy Whitfield Marketing Science 2019 Vol. 38 No 6, pp. 918-926.

III. EXTERNAL SOURCES OF CONTROL: PRIVACY REGULATION AND COPYRIGHT AS SOURCE OF CONTROL OVER DATA

The previous section discussed limitations in terms of imposing internal control over data. However, I would argue that there are still concerns regarding external sources of control. I would argue that this is most likely to be an issue when the data in question is also potentially governed by privacy or copyright protections.

A useful example where data segmentation services are treated as though it was an essential facility is that of IMS Health and NSC Health corporations. These are two competitors in pharmaceutical data services in Germany which provided sales reports from individual pharmacies.

Due to German privacy laws, data has to be appropriately anonymized in a way which is privacy-compliant. The most practical way of doing this (without losing all marketing insight from the data) is to aggregate it to the postcode level. This was done by something referred to as the "brick structure," which grouped together pharmacies into commercially useful geographical clusters which would not permit the identification of any one pharmacy and complied with the German privacy rule that at least 3 pharmacies had to be aggregated. However, IMS asserted copyright over the brick structure. Any other means of aggregation were potentially not privacy-compliant, because differential aggregation could in theory then be used to identify an individual pharmacy if a firm cross-referenced the two datasets. At this point, the European Commission ordered IMS to grant access to the brick structure on commercially reasonable terms.

It is notable though that in this case it is not data *per se* that is the essential facility but instead the segmentation and parsing of the data.[13]

With the growing presence and salience of privacy regulation such as GDPR and the California Consumer Privacy Act, however, it seems likely that privacy regulation may allow data to become more excludable. In Miller & Tucker (2014),[14] we discuss how hospitals often limit the portability of patient data or intentionally "silo" data while citing privacy concerns as their motivation for doing so. However, we show that the likelihood of siloing data under the guise of privacy regulation often seems to reflect the competitive structure of the local market. I modeled this as well in Campbell et al. (2015),[15] where we showed that the more strenuous the opt-in requirements are, the more likely it is that consumers will opt into sharing their data with just a few firms, rather than being willing to share data with new firms or potential entrants.

As I point out in a new paper,[16] in the past five years there appears to be evidence that in some industries where privacy concerns are particularly salient, concentration of firms increases. One industry I highlight where this has occurred is that of educational technology. The complex nature of student privacy has potentially facilitated two firms (Apple and Google) being able to supply a great deal of educational software to K-12 schools, because school administrators feel reassured that they are sufficiently privacy-compliant compared to educational technology startups in this space.

13 https://ec.europa.eu/competition/antitrust/cases/dec_docs/38044/38044_15_5.pdf.

14 Miller, A. & Tucker C., (2014, January). Health information exchange, system size and information silos. Journal of Health Economics 33 (2), 28.

15 Campbell, J., Goldfarb, A. & Tucker, C. (2015), "Privacy regulation and market structure," Journal of Economics & Management Strategy 24 (1), 47.

16 See https://www.brookings.edu/wp-content/uploads/2019/12/ES-12.04.19-Marthews-Tucker.pdf.

CPI COMPETITION POLICY INTERNATIONAL

CPI Subscriptions

CPI reaches more than 35,000 readers in over 150 countries every day. Our online library houses over 23,000 papers, articles and interviews.

Visit competitionpolicyinternational.com today to see our available plans and join CPI's global community of antitrust experts.