APRIL 2021

DEVOTED TO LEADERS IN THE INTELLECTUAL PROPERTY AND ENTERTAINMENT COMMUNITY

VOLUME 41 • NUMBER 4

THE Licensing Journal®

Edited by Gregory J. Battersby and Charles W. Grimes

Wolters Kluwer

The Data Protection and Monetization Playbook: Is Your Company Ready?

Efrat Kasznik

Efrat Kasznik is president of Foresight Valuation Group, a Silicon-valley based consulting firm specializing in IP valuation and strategy advisory. Ms. Kasznik is a Lecturer on IP Management at the Stanford Graduate School of Business. She serves as Chair of the Licensing Executives Society (LES) USA-Canada Valuation & Pricing Committee, and has been listed on the IAM Strategy 300 - The World's Leading IP Strategists list every year since 2013.

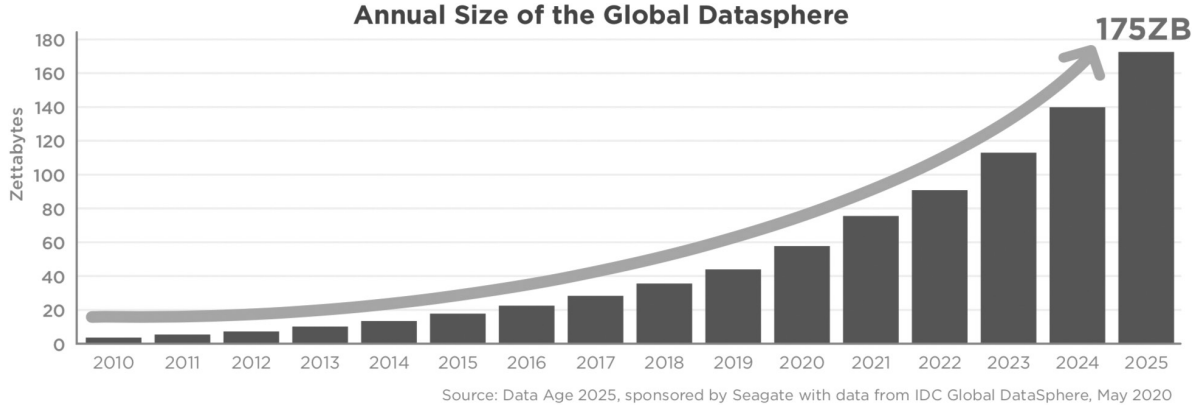
Data Assets Are Key Resources in the Global Economy

The 21st century economy relies on data, which is a new type of intangible asset that can be viewed as the ***digital intangible*** fueling technology companies in all sectors, from banking to manufacturing to biotech. In 2017, The Economist has declared that "the world's most valuable resource is no longer oil, but data".[1] "Data" can be construed to imply a wide variety of compilations of information, but in the context of this article we will refer to the *digital, readable, machine-accessible format of data*. In Data Age 2025 (May 2020), research firm IDC has defined three primary locations where data is created and located: the core (traditional and cloud data centers), the edge (enterprise infrastructure and branch offices), and the endpoints (PCs, smart phones, and IoT devices). IDC predicts that the *Global Datasphere* (defined as all data created, replicated or stored in the above three locations) will grow almost 4 times, from 45 Zettabytes (ZB) in 2019 to 175 ZB by 2025 (1 ZB = 1 trillion Gigabytes), as seen in *Figure 1*.

The digitization of information facilitates the query and analysis of large quantities of data, which enable new business models for monetization. Networking and cloud technologies allow the transfer and storage of large amounts of data with easy access for analysis, and Artificial Intelligence (AI) is changing the way data is interpreted for business decisions. Companies serving the cloud are seeing skyrocketing growth and valuations; just this week (on September 16, 2020), Snowflake—a cloud data-warehousing company (an area of services that did not even exist a decade ago)—had what is described as the largest IPO of a software company ever, having raised about $3 billion (based on opening day prices) at a valuation of over $70 billion.[2]

That being said, the emergence of the Internet of Things (IoT) and other decentralized ecosystems for data collection through networks of sensors and devices, creates challenges around the ownership, privacy and protection of data. A new paradigm viewing

Figure 1.



the enterprise as the "steward of data" imposes obligations and regulations related to the prudent way of collecting and leveraging data.

Data Assets Are Largely Protected as Trade Secrets

It is important to keep in mind that Data assets are not protectable by patents. The main IP protection afforded Data assets is trade secrets protection, which is generally implemented through mechanisms such as strict authentication measures around the access to data, cybersecurity defenses warding off cyber-attacks, and strongly-worded legal contracts governing data access from both inside and outside the organization.

According to Gartner estimates,[3] worldwide cybersecurity spending in 2017—2019 has exceeded (or is expected to exceed) $100 billion annually (as seen in *Table 1):*

A Gartner study (2018)[4] further projects that the cybersecurity market size will increase to $270 billion by 2026. Much of the growth in cybersecurity spending emerges from the high economic cost of data breaches (a risk that is only expected to intensify in the post COVID-19 environment, due to the increase in remote workforce): according to an IBM survey (2019),[5] the average cost of a data breach in the U.S. has more than doubled from $3.54 million in 2006 to $8.19 million in 2019.

Data as a Key Driver of Corporate Value

Data asset have been growing in significance as one of the key drivers of corporate value, particularly in software companies where digital information is more easily generated though users, and gradually in many other types of companies along with the proliferation of IoT ecosystems in many industries. Software Unicorn (pre-exit startups with valuations exceeding $1 billion) valuations are an example of the data-centric valuations that started showing up in the late-90s, along with the dot.com era in the early days of the Internet. Large funding rounds and acquisitions of pre-revenue companies have gradually become more common, particularly when it comes to software companies in business-to-consumer (B2C) verticals such as social media.

Table 1.

Worldwide Security Spending by Segment, 2017-2019 (Millions of U.S. Dollars)

Market Segment	2017	2018	2019
Application Security	2,434	2,742	3,003
Cloud Security	185	304	459
Data Security	2,563	3,063	3,524
Identity Access Management	8,823	9,768	10,578
Infrastructure Protection	12,583	14,106	15,337
Integrated Risk Management	3,949	4,347	4,712
Network Security Equipment	10,911	12,427	13,321
Other Information Security Software	1,832	2,079	2,285
Security Services	52,315	58,920	64,237
Consumer Security Software	5,948	6,395	6,661
Total	**101,544**	**114,152**	**124,116**

Source: Gartner (August 2018)

With the advent of smartphones in the mid-to-late 2000s, customer acquisition became relatively easy, particularly with mobile apps, the vast majority of which were offered for free download. While advertising has traditionally been the revenue model for B2C software companies, many of them have been opposed to ads for reasons related to product design and consumer preferences, and as a result, have generated little to no revenues while amassing large volumes of users. And yet, despite the absolute lack of revenues or any tangible assets (such as product inventories), some of these companies have exited in valuations ranging in the billions of US dollars.

The key to understanding some of these valuation 'anomalies', which also helps frame the data monetization models that will be presented next, is by viewing users as 'bundles of data'. Take, for example, two of the most prominent data-centric transactions driven by users: the 2012 acquisition of Instagram for $1 billion (with approx. 30 million reported monthly active users) and the 2014 acquisition of WhatsApp for $19 billion (with approx. 450 million reported monthly active users), both pre-revenue startups acquired by Facebook. While there arguably may have been some value in the Technology bucket of these two companies, these were both mobile applications operating on a fairly standard technology platform and it is unlikely that this was the basis for billions of dollars in valuations. The value was really embedded in the users, which are proxies for data and represent future monetization options. Indeed, Facebook went on to realize significant returns on these users.

The Enterprise as a the "Steward of Data"

While users may as well be priced as valuable 'bundles of data', one of the questions hampering the monetization of user data, and other types of data collected by the enterprise, has been: *who owns the data?* This question is particularly challenging in IoT environments. Take, for example, a smart home device such as the Google Nest thermostat. The device is installed in private homes, collects information on ambient temperatures and user heating preferences, and translates that data into heating and cooling controls inside the home via the HVAC system. Since there are multiple parties involved in the process, access to the data can be controlled by one or more of these parties:

- *The end user*—the owner of the home where the Nest thermostat is installed, who allows the collection of data required for the operation of the system;
- *The hardware maker*—Google, who makes the Nest thermostat, and who stores all data in the cloud to then be utilized by AI algorithms to control and improve energy consumption related to heating and cooling;
- *The energy utility*—provides the physical infrastructure for heating and cooling through gas and/or electricity, and collects key data related to actual energy consumption;
- *The solar company*—in case of a solar home, there will also be the solar company (such as SunRun, that in most cases leases the system to the homeowner), and collects data related to energy generation through the solar panels.

The ambiguity surrounding data access and ownership associated with IoT ecosystems and other similar networks, accentuates the role of the **enterprise as a "Steward of Data",** a concept highlighted by IDC in their Data Age 2025 study.[6] With the transition to cloud hosting and management of data, more and more of consumer data is collected and kept by enterprises they do business with. The responsibility to maintain and manage all this consumer and business data supports the growth in cloud hosting through data centers. As a result, the enterprise's role as a data steward continues to grow. In fact, the role of safeguarding consumer data touches on issues of security and privacy that often need to be regulated at the national level. In the healthcare field, the Health Insurance Portability and Accountability Act of 1996 (HIPAA)[7] is one such example of a US federal law that requires the creation of national standards to protect sensitive patient information from being disclosed without the patient's knowledge or consent. HIPAA provisions also mandate the adoption of Federal privacy protections for individually identifiable health information. The HIPPA rules apply to entities such as health plans and health care providers dealing with patient data.

More recent initiatives related to corporate data stewardship include the enactment of the General Data Protection Regulation (GDPR),[8] Europe's landmark data privacy and security law, which went into effect in May 2018. GDPR applies to organizations that process personal data of EU citizens or residents as well as organizations that offer goods and services to EU citizens or residents. Hefty fines are imposed on violators, according to a tiered scale based on the severity of violations. One of the most publicized

provisions of GDPR relates to people's right to erasure known as the "right to be forgotten", which grants individuals the right to ask organizations to delete their personal data. In the US, the California Consumer Privacy Act of 2018 (CCPA),[9] which went into effect on January 1, 2020, gives consumers more control over the personal information that businesses collect about them and secures new privacy rights for California residents.

Data Monetization: Four Leading Business Models

Against this backdrop of exponentially increasing volumes of data collected and processed, on the one hand, with strict data privacy regulations and mounting security threats, on the other hand, data monetization largely remains limited in scope. The State of Dark Data, a survey of 1,300 IT and business leaders conducted by data management platform, Splunk, reveals that 55% of the surveyed organizations' data is "dark", which is defined as "untapped and, often, completely unknown".[10] Yet, the vast majority of survey participants agreed that data is "extremely valuable for success".

The matrix in *Figure 2* presents a novel framework for mapping out the various business models associated with active data monetization. This is a dynamic model, updated frequently based on experience gained through client projects and observations in the market. This framework is based on the type of customer: Business to Business (B2B) or Business to Consumer (B2C), on one dimension; and the monetized party (who is paying for the data): User or Third Party, on the other dimension. Identifying the monetized party is critical, due to the proliferation of three-way monetization schemes. We will discuss in detail the underpinnings of each model, with examples of companies or sectors in the market that have successfully implemented each business model.

SaaS and Advertising Models: The Status Quo

The *Software-as-a-Service (SaaS)* business model, and the *Advertising* business model, are two of the most commonly applied data monetization schemes in the market today:

- ***The Software-as-a-Service (SaaS) model*** is a subscription model common in business-to-business (B2B) situations where the user (a business) is paying for access to software or data. This is a model based on *access fee,* and is the most comparable to patent licensing of all four data monetization models presented here. One example of SaaS access to data is the digital subscription service of LexisNexis, providing online access to case law and other legal information via a monthly subscription. Unlike patent licensing, the data

Figure 2.

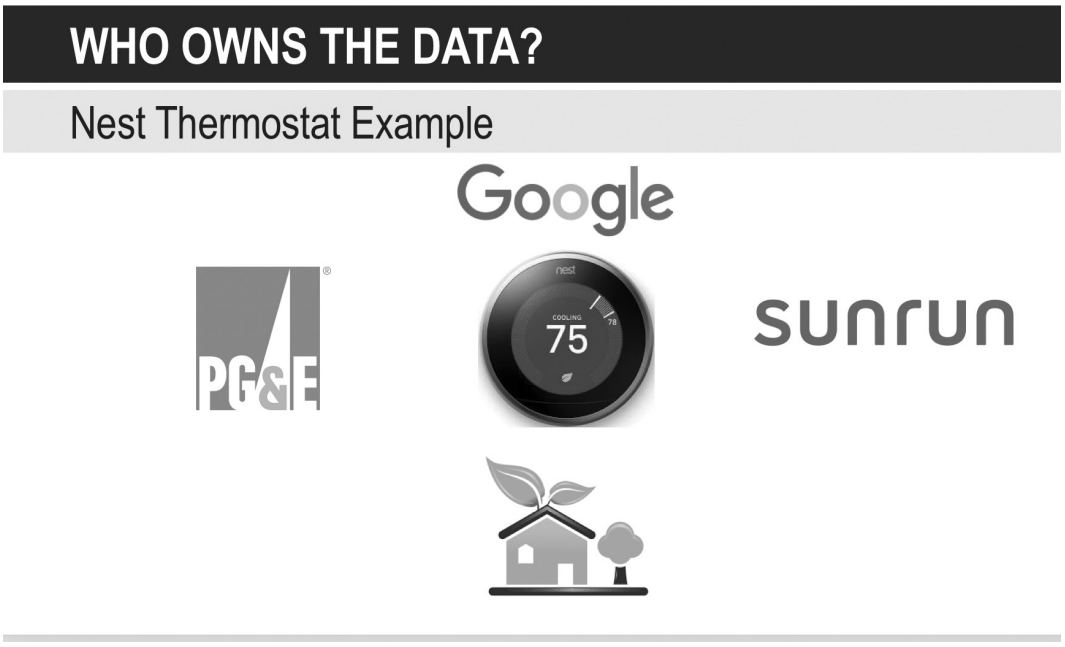
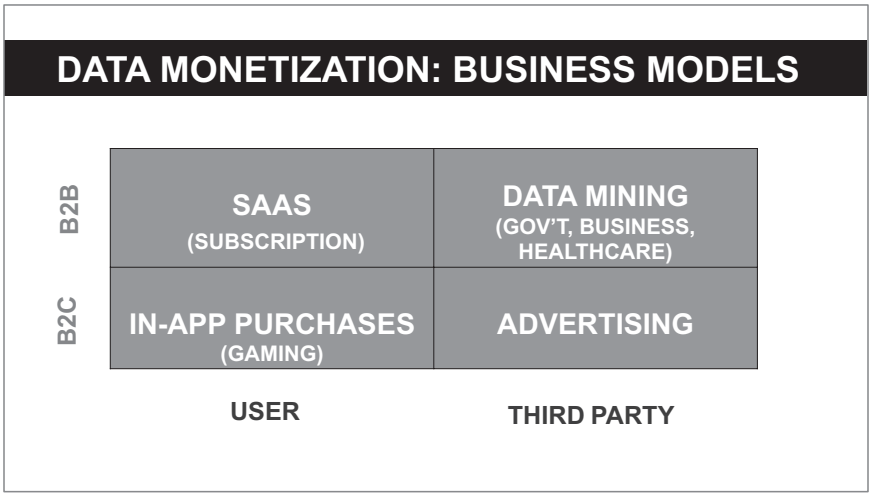


Figure 3.



Source: Foresight Valuation Group, LLC

accessed by LexisNexis subscribers is not proprietary data, but rather aggregated through public sources (some of it may be copyrighted to publishers who used to aggregate it in books, prior to the availability of digital access). The subscription fee for this type of services can be viewed as a "convenience" fee: legal information is becoming voluminous, and paper access is getting impractical. According to its most recent Annual Report, the LexisNexis legal and news database contains 119 billion documents and records including 250 million court dockets and documents.[11] The subscription business model has seen a significant shift over the past 10 years as the legal services market has shifted away from print and now relies on online access.

Monetizing data via a SaaS model is one of the most straightforward business models, and can fit almost every industry where access to large amounts of data is necessary, including such diverse industries as agriculture, transportation and biotech. Its advantages are in the ease of delivery and access, the recurring nature of revenues (metrics like MRR—monthly recurring revenues—are frequently tracked), the ease of "upselling" additional products and offerings to existing customers and the convenience of real-time updates (in the case of LexisNexis, this replaces the need to buy new hard cover editions every year). The main disadvantage of the SaaS business model lies in the risk of *customer churn* (measured by the percentage of existing customers leaving every month). Since *customer acquisition cost (CAC)* is spent upfront in sales and marketing, and revenues are realized in small monthly increments over time, high churn is not a desirable outcome since it reduces the *lifetime value (LTV)* of a customer, a key success metric in SaaS.

- ***The Advertising model*** has been the most common monetization scheme since the dawn of the Internet and is particularly common with business-to-consumers (B2C) applications. This model is based on a *three-way monetization*: consumers access online applications (web or mobile) for free, their data is aggregated and made accessible to third party advertisers, who are the ones essentially covering for the free service via advertising spending on the app or website. The monetized party is not the user, but a third party (advertising brand). In order for this process to work efficiently, there are sophisticated advertising networks serving the ads, and other technology infrastructure that facilitates the matching of customers and messaging. Both Google and Facebook generate most of their revenues from advertising, based on this general three-way model: Google generated over 83% of its $161.9 billion revenues in 2019 from advertising across its various platforms[12]; likewise, Facebook generated $69.7 billion from advertising in 2019, more than 98% of its total revenues for the year.[13]

  Advertising is one of the most ubiquitous data monetization models as almost every free B2C app has some component of advertising revenues supporting its operations. Its advantages are in its simplicity, and self-propelling nature: the ability to offer free access attracts more users who, in

turn, provide more and more data, which attracts more advertisers. While classified as an active monetization scheme, this type of model runs on auto-pilot, as long as there is significant traffic to a site or app. The key disadvantage of this model is that it constantly tests the boundaries of consumer privacy and data stewardship. With the advancement of sensors on mobile devices, and the ability to capture sensitive data like biometric information, privacy concerns intensify and become a target of government intervention which could impede monetization going forward.

It should be noted that, while businesses traditionally pay for SaaS services and consumer data is traditionally monetized through Advertising, we are seeing a convergence through some hybrid business models where consumer SaaS is showing up. This type of hybrid is particularly common in **wearables**, which is a segment of the IoT market. Wearable devices (smart glasses, smart watches, or any other connected device worn on the body that can take vital measurers) collect health data that consumers may be interested in paying access for. The business model for some of these usually includes a free tier of basic access to data, and a paying tier (a model known as "freemium") of access to things like data history over time, data analytics, nutritional recommendations, etc. There may also be an advertising layer on top of that, so these apps also include the three-way monetization that is common for B2C services (albeit at the odds of running into regulatory challenges, which are much higher with health data).

In-App Purchases and Data Mining Models: The Next Frontier

The *In-App Purchases* model and the *Data Mining* model are the more innovative models on the data monetization matrix, and both are still shaping up and evolving in the marketplace. These represent the future of data monetization:

- ***The In-App Purchases model*** is one of the few instances where data monetization is taking place at the consumer level (it's a two-way model, involving the consumer user as the paying party). While consumers do not like to pay for downloading apps (as Apple CEO Tim Cook recently testified, 84% of apps on the Apple App Store are free apps), gaming apps are an exception to the rule. In some games, gamers can pay for digital currency that allows them to buy accessories in the game, also known as "game cosmetics" (modifiers that change the way certain objects look in the game). Gaming is used here as an example of the types of models involving the monetization of digital assets, an extension of data into other digital commodities which form a new class of *digital intangibles*. These models involve the use of digital assets either as payment mechanisms, such as tokens (common currencies in blockchain decentralized networks), or as the goods being acquired in virtual environments (such as in gaming apps). In-App purchases made headlines when Epic Games, publisher of the hugely popular game Fortnite (which has allegedly been downloaded on the Apple App Store nearly 130 million times)[14] announced in August 2020 that a new direct payment option for players is available to purchase the currency used in the game outside of the iOS App Store or Google Play.[15] This direct payment option cut Apple and Google from their revenue share (30% of all app related revenues) as the transaction would not go through their respective platforms. In response, Apple and Google pulled the app from their app stores for violations of their terms of service, and Epic subsequently filed suit against both companies alleging antitrust violations.[16]

  The legal battle surrounding Fortnite shows both the pros and cons of the In-App Purchasing model. On the pros side, this model has appealing economics, as it provides revenues from sales of virtual goods with no cost to create or deliver. Consumers will not pay for the app, but they will pay for the virtual goods, so it taps into consumer behavior in very powerful way. However, the thorn in the otherwise appealing profit margin opportunity is the high cost of the carrying platform, as embodied in the 30% charged by Apple, which gave rise to Epic's legal battle. It is interesting to watch how this situation gets sorted out, as the market will need to find an equilibrium that works for both sides; the role of the platforms is critical in distributing the game, but at the same time, 30% of revenues may be a bit steep for the game publishers.
- Finally, the most ambitious model on this map is the ***Data Mining model***, which represents the "holy grail" of data monetization at the corporate level. This is where the market did not quite figure out yet all the possibilities, as issues of data ownership, security and privacy are major hurdles to fully realizing the potential of data mining. This is a multi-party monetization

model, involving large scale data collected across industries, devices and physical environments. The pioneers in data mining are governments and healthcare systems, who have access to data at a large scale, and use predictive analytics and other tools to drive public health policy (such as in the recent COVID-19 pandemic) or for national security purposes. The scale of data collection and analytics involved here are often beyond the capabilities of most government agencies or corporations, so what emerged in the market are intermediary platforms that process the data and share the results with customers under various arrangements.

One data mining platform that stands out is **Palantir**, which recently filed for an IPO, providing a rare glimpse into its highly secretive operations According to Palantir's prospectus, their software platforms are used by many of the world's most vital institutions, from defense and intelligence agencies to companies in the healthcare, energy, and manufacturing sectors.[17] Palantir offers two software platforms, Palantir Gotham and Palantir Foundry. Gotham was constructed for analysts at defense and intelligence agencies who were "hunting for needles in not one, but in thousands of haystacks." Foundry was built for commercial institutions to create a central operating system for their data. In H1 2020, Palantir's platforms were used by 125 customers, including the U.S. Army. Palantir's prospectus provides only a hint of the benefits derived by their customers, stating that their pricing is based primarily on the expected value that their platforms produce for their customers. Other companies providing similar services include: Tableau, Cloudera, Teradata, and Qlik.

Conclusion

The role of the organization as a "steward of data" should not be misconstrued as an injunction on data monetization. On the contrary: data has the potential to enhance corporate value in significant ways, and data assets should be viewed as an integral part of the modern IP portfolio (as *digital intangibles*). Using the analogy of data as the new fuel, the engines of data analytics are already revving up, we just need to properly address roadblocks such as privacy and security. Every technology company that wants to jump on the data monetization train should be very familiar with the types of models and platforms allowing it to leverage its data assets in either a two-way or a three-way model, and eventually embark on the full benefits of data mining when the time is right.

A modified version of this article first appeared in IAM Magazine.

1. *See https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data.*
2. *See https://www.economist.com/business/2020/09/15/how-snowflake-raised-3bn-in-a-record-software-ipo.*
3. *See https://www.gartner.com/en/newsroom/press-releases/2018-08-15-gartner-forecasts-worldwide-information-security-spending-to-exceed-124-billion-in-2019.*
4. *See https://www.forbes.com/sites/louiscolumbus/2020/04/05/2020-roundup-of-cybersecurity-forecasts-and-market-estimates/?sh=722f367e381d.*
5. *See* "Cost of a Data Breach Report" *https://www.ibm.com/downloads/cas/ZBZLY7KL?_ga=2.181209767.1686129232.1586105132-321662522.1584074488.*
6. *See* "The Digitization of the World from Edge to Core" *https://www.seagate.com/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf.*
7. *See https://aspe.hhs.gov/report/health-insurance-portability-and-accountability-act-1996.*
8. *See https://gdpr.eu.*
9. *See https://oag.ca.gov/privacy/ccpa.*
10. *See https://www.splunk.com/en_us/form/the-state-of-dark-data.html.*
11. *See https://www.relx.com/~/media/Files/R/RELX-Group/documents/reports/annual-reports/2019-annual-report.pdf.*
12. *See https://www.sec.gov/Archives/edgar/data/1652044/000165204420000008/goog10-k2019.htm.*
13. *See http://d18rn0p25nwr6d.cloudfront.net/CIK-0001326801/45290cc0-656d-4a88-a2f3-147c8de86506.pdf.*
14. *See https://www.documentcloud.org/documents/7203851-Epic-v-Apple-counterclaims.html.*
15. *See* "The Fortnite Mega Drop-Permanent Discounts Up to 20 Percent" *https://www.epicgames.com/fortnite/en-US/news/the-fortnite-mega-drop-permanent-discounts-up-to-20-percnt.*
16. *See* "The Dean Beat: Apple v. Epic-A Briefing on the Antitrust Arguments and Interesting Facts" *https://venturebeat.com/2020/09/11/the-deanbeat-apple-v-epic-a-briefing-on-the-antitrust-arguments-and-interesting-facts/.*
17. *See https://www.sec.gov/Archives/edgar/data/1321655/000119312520244936/d904406ds1a.htm.*