

MIT Open Access Articles

Data Platforms and Network Effects

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

Citation: Cusumano, Michael. 2022. "Data Platforms and Network Effects."

As Published: <https://doi.org/10.1145/3555833>

Publisher: ACM|Communications of the ACM

Persistent URL: <https://hdl.handle.net/1721.1/146477>

Version: Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

Terms of Use: Article is made available in accordance with the publisher's policy and may be subject to US copyright law. Please refer to the publisher's site for terms of use.





DOI:10.1145/3555833

Michael A. Cusumano

Technology Strategy and Management

Data Platforms and Network Effects

How data-network effects create opportunities and inflate expectations.

INDUSTRY PLATFORMS ARE FOUNDATIONS that bring people and organizations together for a common purpose, which usually includes making money. They function at the level of a market or ecosystem, rather than only within a specific firm. They often start with products such as operating systems and microprocessors, services such as social media and messaging systems, or marketplaces for e-commerce and financial transactions. They can link thousands, millions, or even billions of users and other market actors. But another type of industry platform has recently received attention from consultants such as The Boston Consulting Group as well as investors, entrepreneurs, and policymakers. These platforms center around data. Some have become extremely valuable. How can data inspire industry platforms and what is their potential as businesses?

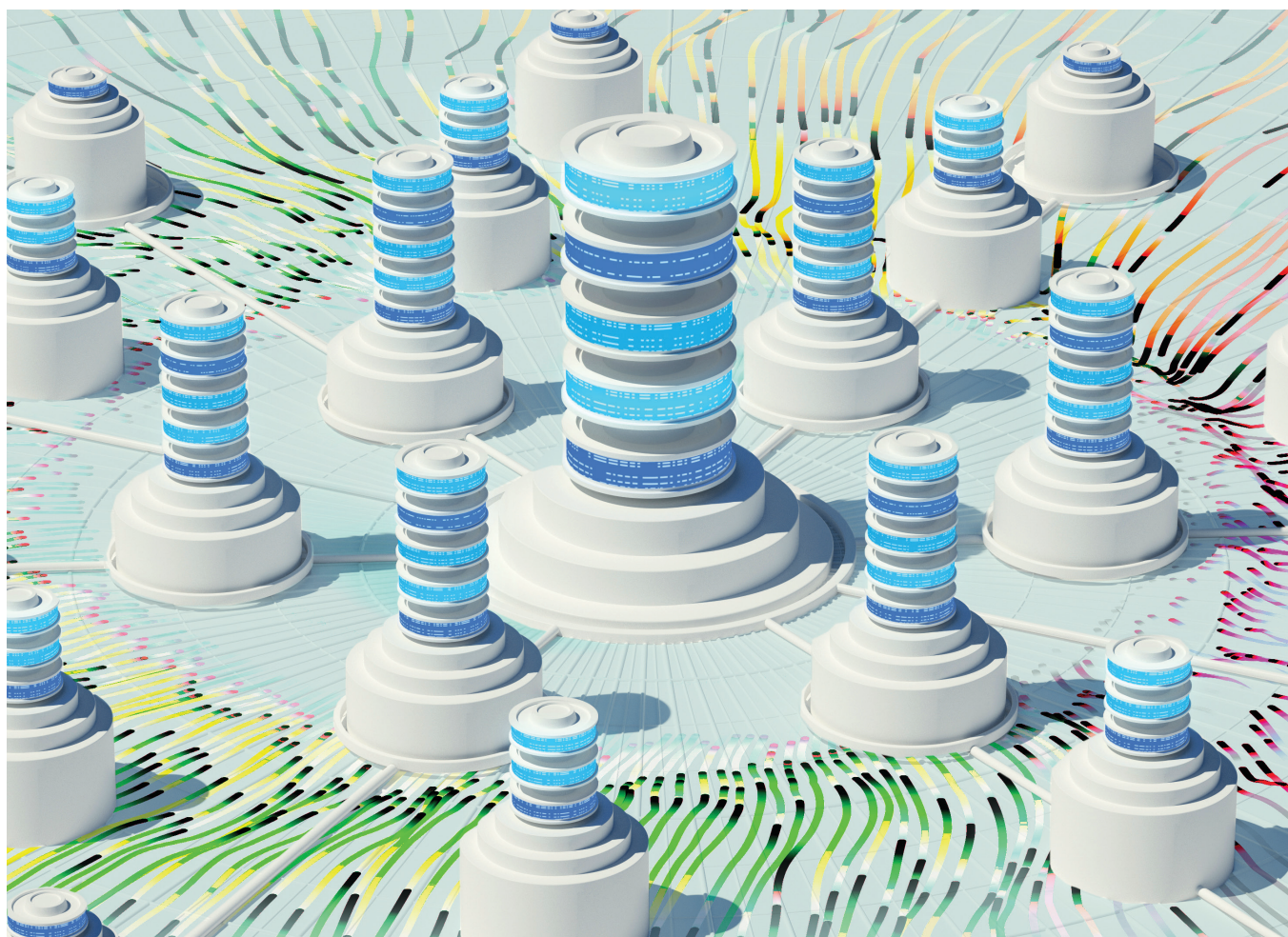
A distinctive feature of industry platforms, and fundamental to their definition, is the ability to generate

positive feedback loops with increasing returns for users and other market participants.⁴ We call these feedback loops “network effects.” They occur on the same-side of the market when a platform connects users directly to other users, such as with the telephone, a social media or messaging

Data can have standalone value as a digital product or service but to evolve into a multisided platform business it needs additional qualities.

app, or any peer-to-peer exchange system. Then cross-side network effects can occur when an industry platform connects demand (for example, buyers) with supply (for example, sellers). Network effects imply the value of the platform increases, at least potentially and sometimes geometrically or exponentially, with each additional user or “complement,” such as apps for a smartphone or drivers for a ride-sharing service. But, when it comes to data, who are the users and what are the demand and supply sides? And where are the network effects, if any?

Data can have standalone value as a digital product or service but to evolve into a multisided platform business it needs additional qualities. First, the data should be in a form suitable for analysis and curation. “Data is the new oil,” as the mathematician Clive Humby stated back in 2006. But most data, like oil, needs refining in order to be useful.⁸ Second, more data should result not simply in more data but in increasingly useful data. Some observ-



ers have called this a “data-network effect.”^{3,13} It resembles a same-side network effect, though here the focus is on the data itself. Third, analyzing the data should reveal useful characteristics of different market actors. These connections provide the basis for cross-side network effects and monetization opportunities. Fourth, as with some other industry platforms, a complementary ecosystem of third-party product and service providers is likely to emerge, powered by cross-side network effects, to help collect, store, analyze, share, and sell the data.¹²

Internet search may be the best-known example of data-network effects that fueled a multisided and multibillion-dollar platform business. The more searches that occur, the more precise the searches become and, at least theoretically, the more each user benefits from improved searches. There may be diminishing returns to scale in search.⁵ Even so, Yahoo, Google, Microsoft, and other firms turned search into a two-sided market by linking advertis-

ers to specific user queries and profiles. TikTok, Facebook, WeChat, Twitter, Amazon, Alibaba, and other social media and ecommerce platforms continually benefit from data-network effects with their recommendation engines as well as advertising sales algorithms, such as in Google AdWords.

A similar platform with powerful data-network effects is for traffic, such as from Waze and Google Maps. The more vehicles tracked by these applications, the more precise the data on traffic patterns and delays, and the better the information provided to drivers, such as time estimates and routing options. Better data provides more opportunities to sell location-specific ads.

How many other data-platform ventures are out there? We only have to look to find them, and their histories predate the digital age. An old data business is ratings, such as from Dun & Bradstreet (established 1841), Standard & Poor’s (1860), and Moody’s (1900) for financial investments, or Equifax (1899) for personal credit

scores. Another is Nielsen, founded in 1923 as a market research firm and now famous for evaluating television shows. A more recent example is BitSight (2011, current market value \$2.4 billion), which created a rating, similar to a credit score, for levels of cybersecurity.¹⁴ Some elements of the scoring system improve with more users, network information and problems identified (viruses or other malware), and data points (computing devices and organizations identifiable on the Internet). BitSight sells the ratings and analyses not only to clients who want to know their vulnerabilities. It also sells to banks, insurance companies, governments, and companies that want to know the cybersecurity risks they face when offering loans and insurance policies or providing suppliers access to their information networks.^a

^a Full disclosure: Two of my former MIT students—Nagarjuna Venna and Stephen Boyer—started BitSight and I serve as ombudsman for their customer council.

Several other data ventures focus on specific markets or technology areas.⁶ Flatiron Health (2012), acquired by Roche in 2018 for \$1.9 billion, aggregates data from oncology clinics, hospitals, and medical research centers, and then sells curated datasets to clinics, researchers, and biotech companies interested in accelerating drug development or improving new treatments. Arity (2016), a subsidiary of Allstate Insurance, analyzes data from more than 20 million drivers and sells it to insurance companies, auto companies, and mobile-app companies. Skywise (2017), launched by Airbus and several partners, helps more than 100 airlines manage aircraft supply, maintenance, and fuel usage for the \$5 trillion aircraft services business. Kabbage (2019), acquired by American Express in 2020 for \$850 million, analyzes real-time data on small businesses, such as from accounting software, bank accounts, and UPS shipping records, to assess credit risk and make lending decisions in minutes rather than days or weeks.

Over the past decade, there has been a kind of “Gold Rush” mentality as startups and established companies competed to find value in big data or the Internet of Things, and then sell the information.¹⁰ We have now entered a period of more realistic valuations. However, as in the historical Gold Rushes, even when the nuggets of treasure prove difficult to find, ecosystem players still can make lots of money selling tools and infrastructure—the digital equivalent of “picks and shovels.”

Not surprisingly, all the major cloud companies now promote technology to facilitate data aggregation, analysis, and sharing or selling.⁹ In 2018, Microsoft, Oracle, and SAP launched their Open Data Initiative to enable data interoperability among their applications.⁷ In 2019, Amazon launched the AWS Data Exchange and Open Data Sponsorship Program to help companies sell data and host public datasets free of charge.^{1,2} Meanwhile, other firms have emerged as leaders in data-infrastructure products and services. Most notably, Palantir (2003, approximately \$17 billion market value) sells operating systems and other tools that enable its customers, largely from different U.S. government branches, to integrate

Some data businesses have retained their multibillion-dollar valuations while others generate value in other ways.

their datasets, analytics, and intelligence operations. Snowflake (2012, approximately \$48 billion) offers a cloud-based data warehouse for customers who do not want to store their data on premise. It also offers SQL tools and a marketplace that resells third-party datasets.

Some data startups have dropped dramatically in value as expectations faded. Metromile (2011), valued at \$2 billion in November 2021 and then sold to Lemonade in July 2022 for \$137 million, sells usage-based insurance to drivers at Uber and other ride-sharing companies. It provides feedback on how to improve fuel usage but Uber continues to lose billions of dollars for other reasons (see “‘Platformizing’ a Bad Business Does Not Make It a Good Business,” *Communications*, Jan. 2020). Otonomo (2015), valued at \$1.4 billion in February 2021 and now worth approximately \$67 million, aggregates and analyzes data for autonomous-vehicle machine learning as well as related insurance, mapping, parking, and other mobile applications. This market has been very slow to evolve (see “Self-Driving Vehicle Technology: Progress and Promises,” *Communications*, Oct. 2020).

Some data businesses have retained their multibillion-dollar valuations while others generate value in other ways. The Boston Consulting Group has identified more than 200 platforms focused on public health, disaster management, the environment, personal mobility and smart cities, agriculture and food supply chains, natural resource management, education, and economic development.¹² For example,

the United Nations’ Humanitarian Data Exchange enables governments to share information for disaster relief and pandemics, among other applications. The World Resources Institute’s Global Forest Watch provides satellite images and data analysis aggregated from various partners to track deforestation. Indigo Ag has a platform to help farmers monetize carbon removal and sequestering on their farmland. Climate TRACE, a global coalition launched in 2021 by former U.S. Vice President Al Gore and others, tracks greenhouse gas emissions.

The goal of “social” data platforms is to illuminate global problems and identify solutions. Their cross-side network effects have the potential to bring individuals and organizations together, not to make money but to tackle critical challenges such as climate change, depletion of environmental resources, and ongoing hunger and poverty. We shall see if these platforms can also generate same-side data-network effects and continually improve their utility. **□**

References

1. Amazon. AWS Announces AWS Data Exchange. Company Press Release (Nov. 13, 2019).
2. Amazon. AWS Public Sector Blog; <https://go.aws/3vQq06T>
3. Currier, J. What makes data valuable: The truth about data network effects. NFX.com (Feb. 20, 2020).
4. Cusumano, M., Gawer, A., and Yoffie, D. *The Business of Platforms: Strategy in the Age of Digital Competition, Innovation, and Power*. Harper Business, New York, 2019.
5. Karzit, T. Google’s Varian: Search scale is bogus. CNET.com (Aug. 14, 2009).
6. Koster, A. and von Szczepanski, K. Building a business from data is hard—Here’s how the winners do it. Boston Consulting Group, 2020.
7. Lardinoio, F. Microsoft, SAP, and Adobe take on Salesforce with their new open data initiative for customer data. Techcrunch.com (Sept. 24, 2018).
8. Palmer, M. Data is the new oil. ANA Marketing Maestros. (Nov. 3, 2006); <https://bit.ly/3zNXRyx>
9. Rice, M. What is a data platform? 28 examples in big data you should know. BuiltIn.com, 2022.
10. Russo, M. and Tian, F. The New Tech Tools in Data Sharing. Boston Consulting Group, 2021.
11. Russo, M. and Tian, F. Where is Data Sharing Headed? Boston Consulting Group, 2021.
12. Russo, M. et al. Sharing Data to Address Our Biggest Societal Challenges. Boston Consulting Group, 2021.
13. Turk, M. The Power of Data Network Effects. mattturck.com (Jan. 4, 2016.).
14. Verma, P. Boston’s BitSight raises \$250 million from Moody’s, as ratings firm gauges corporate America’s cyber risk. *Boston Globe* (Sept. 13, 2021).

Michael A. Cusumano (cusumano@mit.edu) is a professor and deputy dean at the MIT Sloan School of Management and coauthor of *The Business of Platforms: Strategy in the Age of Digital Competition, Innovation, and Power* (2019).

My thanks to Massimo Russo of The Boston Consulting Group and Imran Sayeed of MIT Sloan for suggesting this topic and leading an MIT class session on data as a platform.

Copyright held by author.