





COMMENT




<https://doi.org/10.1057/s41599-020-00664-y>

OPEN

Data sovereigns for the world economy

Chunlei Tang ^{1,2}, Joseph M. Plasek¹, Yangyong Zhu³  & Yajun Huang⁴

With the rise of data capital and its instantaneous economic effects, existing data-sharing agreements have become complicated and are insufficient for capitalizing on the full value of the data resource. The challenge is to figure out how to derive benefits from data via the right to data portability. Among these, data ownership issues are complex and currently lack a concept that enables the right to data portability, is conducive to the free flow of cross-border data, and assists in the economic agglomeration of cyberspace. We propose defining the term “data sovereign” as a person or entity with the ability to possess and protect the data. First, the word “sovereign” is borrowed from the fundamental economic notion of William H. Hutt’s “consumer sovereignty.” This notion of sovereignty is strengthened by Max Weber’s classic definition of “power” – the ability to possess any resource. We envision that data capital would provide greater “cross-border” convenience for engaging in transactions and exchanges with very different cultures and societies. In our formulation, data sovereign status is achieved when one both possesses the data and can defend any attack on that data. Using “force” to protect data does not imply an abandonment of data sharing. Rather, it should be easy for an organization to enable the sharing of data and data products internally or with trusted partners. Examples of an attack on the data might be a data breach scandal, identity theft, or data terrorism. In the future, numerous tedious, time-consuming, non-artistry, manual occupational tasks can be replaced by data products that are part of a global data economy.

¹ Brigham and Women’s Hospital, Harvard Medical School, Boston, MA 02120, United States. ² Clinical and Quality Analysis, Mass General Brigham, Boston, MA 02145, United States. ³ School of Computer Science, Fudan University, Shanghai 200438, China. ⁴ School of Economics, Fudan University, Shanghai 200433, China. email: yyzhu@fudan.edu.cn

Problem statement

Data is a resource that has asset value (Fisher, 2009) and can be economically capitalized (Mayer-Schönberger and Ramge, 2018), meanwhile data protection and privacy regulations are rapidly evolving across the globe (European Union, 2016; China, 2017; Monteiro, 2018). Data sharing agreements are increasingly complex yet often insufficient which has in turn resulted in data producers not being able to capitalize on the full value of their data resources (OECD Publishing, 2011). Further, trade in online data-oriented services are facing obstacles at increasing frequency due to political meddling via antitrust probes or antiterrorism measures like the US Patriot Act. Political interventions tend to lead to an undermining of individual self-determinism and autonomy as they disrupt the natural progression of a global data economy. We propose defining the term “data sovereign” [noun] as a person or entity with the ability to possess and protect the data. Here, the word “sovereign” is borrowed from the fundamental economic notion of “consumer sovereignty (Treasury Board of Canada Secretariat, 2006)” that William H. Hutt coined to denote “... the market economy, the production decisions of entrepreneurs are rigidly governed by the freely expressed spending decisions of consumers (Salerno, 2009).” The term data sovereign is the needed term to appropriately fill in an existing terminology gap as it, for example, enables rights to data portability, is conducive to the free flow of cross-border data, and assists in the economic agglomeration of cyberspace.

Data ownership is initially assigned by default to the person providing the data (e.g., a social media user) to a particular entity (e.g., organization, State, country) (Fig. 1). Just like Zuboff enumerated in her *Surveillance Capitalism* (Zuboff, 2019): “The idea of ‘data ownership’ is often championed as a solution. But what is the point of owning data that should not exist in the first place? It’s like negotiating how many hours a day a seven-year-old should be allowed to work, rather than contesting the fundamental legitimacy of child labor.” In contrast, a data producer is any agent who has the ability to add value to raw data or other primary data products rather than the individual providing their personal data. Existing terms, such as data residency and data localization, that describe the global flow of data were developed for cloud infrastructure technologies. Data residency refers to the geographic locale in which the organization physically stores its data. Data localization differs from data residency in that data is additionally subject to the laws of the country in which it is physically stored. In lay use, data residency and data localization are used so interchangeably that their individual meanings have become lost and indistinguishable.

Proposed concept

In contrast to the above terms, we propose that the data sovereign is initially appointed as the entity possessing the data (e.g., the social media platform, government entity). In our formulation, data sovereign status is achieved when one both possesses the data and can defend any attack on that data. Examples of an attack on the data might be a data breach scandal (Snider and Baig, 2019), or an infiltration by a State actor. Weber defines “power” as the ability to possess any resource (Trans. Waters, Waters, et al., 2010). Following Weber’s formulations, data as an economic resource should be preserved through “the sole grantor of the right to physical force (Waters, 2015)” with the legitimacy of domination. Such a willingness of enterprises to convey power may effectively align incentives for data innovations towards enhancing data resource protection capabilities. Using “force” to protect data does not imply an abandonment of data sharing. Rather, it should be easy for an organization (e.g., corporations, industry bodies, government entities) to enable sharing of data and data products internally or with trusted partners. From an operational standpoint, a data force belongs to a specific data sovereign or an alliance of data sovereigns and functions like a guild of mercenaries who conduct data activities. The core function of the data force is the defense and exploitation of the data sovereign’s data and data products. In essence, the data force aims to enhance data and data products by repairing or mitigating vulnerabilities through the use of data innovations. Wargaming, the simulation of different confrontations, is a strategy used by data forces to identify and patch vulnerabilities, and may include simulation of attacks on other data sovereigns as pre-emptive strikes.

The results of a COVID-19 screening test are owned by the patient, who often takes possession of them through a patient portal into the electronic health record system or via a phone call from the healthcare facility. These data are of use to public health officials for contact tracing of positive cases to identify those whom the patient came in contact with while they were contagious. The lab is (likely) mandated by regulations to report positive cases to the local public health office, who then becomes the data sovereign as they are the ones compiling statistics from multiple labs in the locale and acting on the data. Data sharing agreements with the Centers for Disease Control may exist with some State and local governments, for example, where their assistance has been requested. Healthcare organizations, and public health officials may work together to create models that simulate the spread of the pandemic, which is useful for resource planning and health policy purposes. Similarly, the fast-growing global online food delivery services market, of which the most significant segment is Restaurant-to-Consumer Delivery, tracks

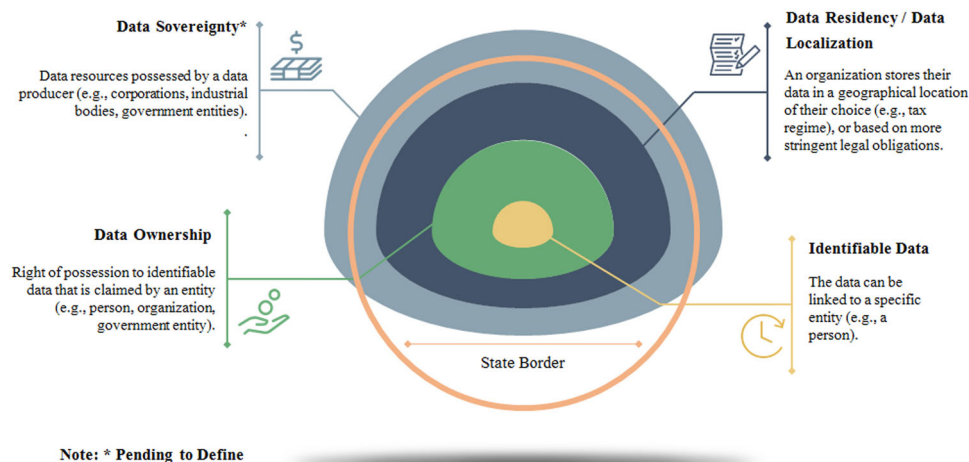


Fig. 1 Lexicon of terms relevant to data disputes.

the food preferences and habits of consumers to improve service, expand the customer base through targeted advertising, and optimize menus to fit the food preferences of the community. Delivery orders, owned by the consumer, are placed through a mobile app (e.g., Uber Eats). This data can be later used to shorten the distance between farm and table, which is particularly helpful to dairy farmers (Waters et al., 2010) in the COVID-19 era. The person must provide a phone number (mobile) and geolocation to the app when buying food for delivery, resulting in the appointment of the data producer (i.e., organization) controlling the app as a data sovereign over the shared data. Local, State, and/or Federal regulations govern how personal information must be protected and how/when it can be used by the organization. For example, some data may require data localization to prevent the data from being stored or transferred internationally, and the ability of law enforcement to utilize geolocation tracking of cell phones varies by locale in terms of privacy protections and warrant requirements. There are legitimate reasons to keep personal data private and to not allow it to be part of the global data economy. A malicious actor such as an identity thief cannot be a data sovereign as illegitimate make money fast schemes require the divestment of collected data or operationalization of the collected data in an attack on legitimate data sovereigns. Similarly, data that is of national security interest has a State or Federal agency as its data sovereign. These data sovereigns must defend this sensitive data from data piracy and data terrorism threats. Social media networks and adult websites are often criticized for profiting off of revenge porn, deep fakes (i.e., AI-generated videos that realistically places new faces onto existing faces in the videos), or content published without a person's/the copyright holder's consent or permission. While these organizations do remove videos with content that violates their terms of service when requested or to comply with a court order, it becomes a game of whack-a-mole unless data tracking technology is put in place at content upload to self-delete content previously flagged or removed. YouTube's Content ID (Shinn, 2015) is an example of such data tracking technology, as it creates a digital fingerprint to allow copyright holders to identify and manage their copyrighted content by blocking new videos, monetizing those videos, or tracking the video's viewership statistics. An alternate strategy to fend off data attacks would be to apply Snapchat's idea to vanish data in the source after transference.

Economic benefits

Finding a term for dedicated use might help resolve a long-standing debate over various data disputes (Fig. 1), such as in where and how to set data resource boundaries in cyberspace (Kalir and Maxwell, 2002). Our goal in coining "data sovereign" is to eliminate some misunderstandings regarding data resources as rightfully belonging to data producers instead of the person providing the data or the State. Regulations in certain countries assign data ownership to the State, which has the potential to significantly disrupt global "data" economic paths to prosperity. For example (Bogaerts and Segers, 2018).

Russia's On Personal Data Law (OPD-Law) requires the storage, update, and retrieval of data on its citizens to be limited to data center resources within the Russian Federation. While laws like this do tend to enhance citizen's privacy from other nation States, they tend to be motivated by national protectionism and contribute to the creation of border-based data silos that obstruct businesses and governments from realizing the full potential of global cross-border data flows. In negotiating a trade agreement, the United States would prefer that (The United States Trade Representative, 2019) the European Union "does not impose measures (e.g., customs duties) that restrict cross-border data flows" and "does not require the use or installation of local computing facilities." The United States takes a different tact with China, blocking both TikTok and a US\$1.2

billion deal between MoneyGram and Ant Financial (a business conglomerate owned by Alibaba Express) (StraitsTimes.Com, 2018) in order to keep personal information and identifying data out of the hands of a foreign entity that they think shares information freely with a foreign government. Our definition stands contrary to digital factionalism and the "splinternet" (Box, 2019) and may thus cause controversy in both academic and industrial circles. We recommend using the power of decentralized markets to protect data via non-legislative acts—production and protection from attack.

When it comes to data, "accessing data trumps owning it (Ogilvy, 2017)," thus data sharing and portability rights are often the main commodity sold or traded to external entities. Like the transfer of State sovereignty in international relations (Taylor, 1997), utilizing the concept of a data sovereign enables the transfer of rights for data sharing and/or portability purposes. The data sovereign can transfer partial rights to governments or other industrial competitors through a modest negotiation. Such a transfer is proactive and strategic when developed by anticipating or analyzing trends and reviewing the organization's past performance concerning externally induced crises or threats. When it comes to data resources, it is often too late to execute a reactive transfer when opportunities arise as these defensive transfers often incur from or result in a considerable loss of data asset valuation.

The Internet Archive was founded by Brewster Kahle to preserve large quantities of the World Wide Web. A recent attack on the Internet Archive was conducted by the US National Writers Union and several digital publishing companies (Hasbrouck, 2020). They listed five distribution channels that indicate how authors (and publishers) were harmed by the presence of an archive, including: (1) downloads via OpenLibrary.org of e-books assembled from page images, (2) audiobooks generated from images of scanned pages, (3) viewing of page images on OpenLibrary.org, (4) viewing of page images on Archive.org, and (5) APIs for automated downloads of page images. Kahle responded to their attack on the Internet Archive by noting that "many of these books are no longer available for sale in the original book form." Books that are no longer available for purchase in their original format don't collect royalties, and thus publishers and authors lack a valid economic argument for harm from the Internet Archive collating this material.

A data sovereign that is inherently concerned with the movement of data may be amenable to the free flow of data across geopolitical borders. We often take for granted that we live in an interconnected world. With the adoption of "data sovereign," boundaries in cyberspace are appropriately defined. For example, Europe's "right to be forgotten" that gives European Union citizens the power to demand data about them be deleted or restricted only applies within the European Union's borders as only this territory is subject to European regulations (unless otherwise negotiated in a trade agreement) (Court of Justice of the European Union, 2019).

The concept of data sovereign enables economic agglomeration in cyberspace. Utilizing data as an economic resource usually has the threat of new entrants and substitute products. When a data sovereign cannot effectively protect their own data resources, it may be prudent to join an alliance of data sovereigns for added protection. This concept of a "data sovereign" may help form data industry clusters in cyberspace that can grow the global data economy through their collaborations. Collective proactive defense and transfer reflects the data sovereign's subjective willingness of using the power of decentralized markets to protect data via non-legislative acts. For example, the co-location of computer servers from multiple organizations in a secure data center provides improved security and economic benefits in terms of rent, personnel, and efficient use of large air conditioning systems creating efficiencies in electricity consumption. Note that such an agglomeration no longer aims to optimize the internal flow of data but

seeks to form a strategic defense in the form of physical and virtual fortifications built by the data sovereign conglomerate.

Conclusion

The history of resource allocation has always been deeply political, and data is now viewed by governments as a valuable resource. An inequity in data-rich profit distribution continues to exist, partially due to government constraints or meddling, which potentially threatens the global data economy and the values of social justice on which it is based. One should be wary of any political determinism or interventions regarding inequalities of data resources exploitation. Instead of anti-trust breakups or mandated takeovers by a non-foreign entity (e.g., Oracle's potential acquisition of TikTok), designating data sovereigns may solve the root problem by relying on the rule of "survival the fittest" that will likely produce smaller tech firms and open the field to more competitors. The approaches of deriving profits from data remain unclear if data producers cannot be confirmed as data sovereigns. Being designated as a data sovereign, to some extent, is an incentive for data producers (e.g., organizations, corporations, industry bodies, or government specifies) to engage in appropriate management of data as a resource. In sum, "data sovereign" is useful from both a data science perspective and an economic perspective. The definition and meaning of "data sovereign" are apposite and may facilitate to solely recognize the sovereign affiliation of a data resource and its possible future capitalization.

Received: 30 June 2020; Accepted: 11 November 2020;

Published online: 16 December 2020

References

- Bogaerts B, Segers K (2018) The "localisation" of Russian citizens' personal data: Compliance with the Russian law on personal data. KPMG Global Sustainability Institute. Retrieved from <https://home.kpmg/be/en/home/insights/2018/09/the-localisation-of-russian-citizens-personal-data.html>
- Box J (2019) Data sovereignty vs data residency vs data localization. Insights For Professionals. Retrieved from <https://www.insightsforprofessionals.com/it/storage/data-sovereignty-data-residency-data-localization>
- China (2017) Information security technology—Personal information security specification. Retrieved from <https://www.tc260.org.cn/upload/2018-01-24/1516799764389090333.pdf>
- Court of Justice of the European Union (2019) The operator of a search engine is not required to carry out a de-referencing on all versions of its search engine. Retrieved from <https://curia.europa.eu/jcms/upload/docs/application/pdf/2019-09/cp190112en.pdf>
- European Union (2016) General Data Protection Regulation. Retrieved from <https://gdpr-info.eu>
- Fisher T (2009) The data asset: how smart companies govern their data for business success. John Wiley & Sons, New York
- Hasbrouck E (2020) What is the Internet Archive doing with our books? Nwu.Org. Retrieved from <https://nwu.org/what-is-the-internet-archive-doing-with-our-books>
- Hutt WH (1940) The concept of consumers' sovereignty. *Econ J* 50(197):66–77
- Kalir E, Maxwell EE (2002) Rethinking boundaries in cyberspace: A report of the Aspen Institute internet policy project. Retrieved from <https://assets.aspeninstitute.org/content/uploads/files/content/docs/cands/RETHINKCYBERSPACE.PDF>
- Mayer-Schönberger V, Rameg T (2018) Reinventing capitalism in the age of big data. Basic Books, New York
- Monteiro RL (2018) The new Brazilian General Data Protection Law – A detailed analysis. International Association of Privacy Professionals. Retrieved from <https://iapp.org/news/a/the-new-brazilian-general-data-protection-law-a-detailed-analysis>
- OECD (Organisation for Economic Co-operation and Development) Publishing (2011) Regulatory policy and governance: Supporting economic growth and serving the public interest. OECD Publishing, Paris
- Ogilvy J (2017) The future according to "Wired" editor Kevin Kelly. *Forbes*. Retrieved from <https://www.forbes.com/sites/stratfor/2017/03/30/the-future-according-to-kevin-kelly/#46b07cbf76bf>
- Salerno JT (2009) The essence of Hutt. Mises Institute. Retrieved from <https://mises.org/library/essence-hutt>
- Schneider L (2020) Dairy farmers dumping milk amid COVID-19: Pandemic's impact on the dairy industry. *abcNews.go.com*. Retrieved from <https://abcnews.go.com/US/dairy-farmers-dumping-milk-amid-covid-19-pandemics/story?id=70268302>
- Shinn LD (2015) Youtube's content ID as a case study of private copyright enforcement systems. *AIPLA QJ* 43:359
- Snider M, Baig EC (2019) Facebook fined \$5 billion by FTC, must update and adopt new privacy, security measures. *USA Today*. Retrieved from <https://www.usatoday.com/story/tech/news/2019/07/24/facebook-pay-record-5-billion-fine-u-s-privacy-violations/1812499001>
- States-European Union negotiations (2019) Retrieved from https://ustr.gov/sites/default/files/01.11.2019_Summary_of_U.S.-EU_Negotiating_Objectives.pdf
- StraitsTimes.Com (2018) MoneyGram's sale to Jack Ma's Ant Financial becomes latest deal torpedoed by Trump administration. Retrieved from <https://www.straitstimes.com/business/companies-markets/moneygrams-sale-to-jack-mas-ant-financial-becomes-latest-deal-torpedoed>
- Taylor CR (1997) A modest proposal: statehood and sovereignty in a global age. *U Pa J Int'l Econ L* 18(3):745–809
- The United States Trade Representative (2019) Summary of specific negotiating objectives on United States-European Union negotiations. Retrieved from https://ustr.gov/sites/default/files/01.11.2019_Summary_of_U.S.-EU_Negotiating_Objectives.pdf
- Trans. Waters D, Waters T, Hahnke E, Lippke M, Ludwig-Glück E, Mai D, Ritzi-Messner N, Veldhoen C, Fassnacht L (2010) The distribution of power within the community: classes, stände, parties by Max Weber. *J Class Sociol* 10:137–152
- Treasury Board of Canada Secretariat (2006) Frequently Asked Questions: USA PATRIOT ACT Comprehensive Assessment Results. Retrieved from https://www.tbs-sct.gc.ca/pubs_pol/gospubs/TBM_128/usapa/faq-eng.asp
- Waters T (2015) Weber's rationalism and modern society: new translations on politics, bureaucracy, and social stratification. Palgrave Macmillan, London
- Zuboff S (2019) The age of surveillance capitalism: the fight for a human future at the new frontier of power. PublicAffairs, New York

Acknowledgements

The authors thank David W. Bates, MD and Sheng Wang, PhD, for valuable comments and suggestions on the early versions. The content is solely the responsibility of the authors.

Author contributions

CT and YZ built on and extended the initial idea. CT drafted the manuscript. All authors provided substantial contribution to the conception and helped revise the manuscript. All the authors are accountable for the integrity of the work.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1057/s41599-020-00664-y>.

Correspondence and requests for materials should be addressed to Y.Z.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing,

adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020