INTERNATIONAL MONETARY FUND

Toward a Global Approach to Data in the Digital Age

Vikram Haksar, Yan Carrière-Swallow,
Andrew Giddings, Emran Islam, Kathleen Kao,
Emanuel Kopp, and Gabriel Quirós-Romero

SDN/2021/005

STAFF DISCUSSION NOTE

2021
October

Toward a Global Approach to Data in the Digital Age

SDN/2021/005

Prepared by Vikram Haksar, Yan Carrière-Swallow, Andrew Giddings, Emran Islam, Kathleen Kao, Emanuel Kopp, and Gabriel Quirós-Romero[1]

ISBN:	9781513599427
JEL Classification Numbers:	D40, D80, E20, F10, F20, F30, F50, F60, G10, G40, G50, H20, K10, L10, L40, L50, O30, O40
Keywords:	Data, finance, bigtech, competition, privacy, trade, policy coordination, global principles
Authors' E-mail Address:	vhaksar@imf.org, ycswallow@imf.org, agiddings@imf.org, eislam@imf.org, kkao@imf.org, ekopp@imf.org, gquiros@imf.org

[1] We are very grateful for guidance from Aditya Narain, Yan Liu, and Martin Cihak; for substantial contributions from Barend Jansen and Ghiath Shabsigh; and for valuable comments from Maria Soledad Martinez Peria, Nicola Pierri, Florian Gimbel, and other colleagues from IMF departments. We are also thankful to Pavel Lukyantsau, Ashley Abraham, and Lilly Siblesz de Doldan for research and production support.

CONTENTS

EXECUTIVE SUMMARY

Ongoing economic and financial digitalization is making individual data a key input and source of value for companies across sectors, from Big Tech and pharmaceuticals to manufacturers and financial services providers. Data on human behavior and choices—our "likes," purchase patterns, locations, social activities, biometrics, and financing choices—are being generated, collected, stored, and processed at an unprecedented scale.

Use of individual data and digital innovation can power productivity; increase access to finance; and promote trade, including of digital services, to the benefit of all. Inspired by the Bali Fintech Agenda (IMF 2018b), this Staff Discussion Note argues that the rules and regulations governing access to this individual data are emerging as a new pillar of structural policies that matter for growth, equity, and financial stability. This imperative is recognized by the G7 Panel on Economic Resilience in its June 2021 statement. The data policy frameworks are being reviewed around the world to strike a balance between privacy and societal needs on one hand and economic and financial benefits on the other, albeit from a mostly national or regional perspective. Key challenges include the following:

- *Balancing privacy trade-offs:* Policies to protect privacy—an important objective in most countries—can mitigate the undesired use of individual data. But protecting privacy can impede private and public efforts to generate economic and social gains from access to data and its use in support of regulatory enforcement and the fight against criminal activity. Clear rules are needed to tackle these trade-offs, including giving people effective control over their data.
- *Promoting inclusive digitalization:* Data can support greater efficiency and inclusion, including in the provision of financial services. But it can also be used for price discrimination and may feed algorithmic biases, disadvantaging and excluding some individuals from important services.
- *Fostering competition in the digital economy:* Individual data as a resource can support productivity, enhancing innovation and the public good—such as for biomedical research. But it can be hoarded by large data collectors, reducing competition, which could dampen innovation and raise financial stability risks.

Domestically, new policy tools and approaches are being considered to provide solutions to these challenges, including mandates for interoperability of networks and data portability, creation of data fiduciaries to help manage consent to protect privacy, and public data utilities to provide data as a public good while protecting privacy. Balancing trade-offs between objectives and the integration of specific policy solutions will require *unprecedented* cooperation among regulators and government agencies with individual mandates for policies and outcomes on competition, financial stability, integrity, consumer protection, and privacy. Promoting biomedical research may call for greater data sharing, but could conflict with privacy. Increasing competition in financial services could expand the perimeter of service providers, challenging integrity and stability.

Internationally, cooperation is critical to contain fragmentation of the global digital economy, which could harm developing economies in particular through the emergence of a digital and data availability divide. Data is the ultimate mobile factor, with cross-border data flows underpinning a rapidly growing proportion of international services trade. But this note argues that the data policies countries are adopting are not always consistent with global welfare. Countries' treatment of privacy, competition, and stability reflects national priorities. The resulting fragmentation could be damaging to smaller countries with smaller data pools and those more dependent on multinational foreign digital firms.

This note argues for development of international agreement on common minimum principles for the data economy. These principles could reduce policy divergences that will arise in the global digital economy, in part reflecting different national contexts and priorities. For example, individual countries may place different emphasis on privacy, but a common understanding of definitions and perimeters of applicability of privacy protection mandates could mitigate avoidable divergences. Many of the other domestic policy approaches being proposed for managing the data economy—for example, on interoperability and data portability to address competition or on the introduction of data fiduciaries to manage consent—could also benefit from common principles on their international application. Likewise, there is scope for international coordination on compilation and sharing of data sources from private digital companies—for regulatory and public policy purposes—including as recognized by the Group of Twenty (G20) ministers in their July 2021 call for a renewed Data Gaps Initiative.

While there is significant uncertainty about how the digital economy will evolve, a global approach will be needed to ensure a level digital playing field for all countries increasingly connected to and dependent on the digital economy. Such an approach will help the digital economy generate value for all without undermining other important macro-financial and social objectives.

Box 1. Key Elements of Common Minimum International Principles

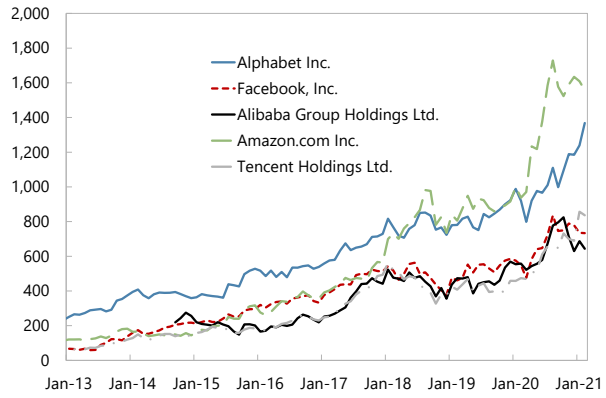
- **Principles for data protection:** An international agreement on common minimum standards for acceptable protection of individual data when it is shared across borders would reduce uncertainty for businesses seeking to comply. Such an approach could draw on the OECD Principles on Privacy (1980 and amended 2013), with further thought on issues such as the role of consent and the definition of data and individuals.
- **Principles on interoperability and data portability:** Given the global reach of businesses that make use of individual data as an important part of their business, there is a need to discuss common principles on how such interoperability and portability should work across borders (Furman and others 2019). Specific use cases include cross-border payments and cross-border data sharing across open banking initiatives (Committee on Payments and Market Infrastructures 2020). A concrete challenge is to coordinate on principles for enabling the interoperability of digital currencies issued by central banks when these can be used across borders, including a method for digitally identifying individuals and standards for digital wallets and data flows.
- **Principles on data sharing for regulatory purposes:** A rigorous data framework should govern not only the protection of data, but also its disclosure to public bodies, including regulatory authorities, where necessary to meet certain public policy objectives. Exemptions to confidentiality and secrecy provisions are already commonplace in many national frameworks (for instance, tax law and anti–money laundering and combating the financing of terrorism) and in line with many international standards and best practices (such as Financial Action Task Force standards and Organisation for Economic Co-operation and Development tax initiatives). Carefully balancing privacy concerns against disclosure for public policy objectives continues to be needed as data regimes evolve. In line with current efforts, and to the extent possible, principles on data disclosure to public authorities should aim to achieve common ground across national frameworks to allow for global data sharing for enforcement or regulatory purposes.

CONTEXT

1. Individual data has become a key input in modern economies and financial systems.

Many of the world's largest firms—in the tech sector but also in many other sectors—now have data on individuals at the core of their business models, and their market value has accelerated during the COVID-19 pandemic (Figure 1). Individual data feeds artificial intelligence algorithms whose predictions power applications from credit provision to investment planning to ad targeting. During the pandemic, there has been a surge in the use of big individual data sets to analyze and mitigate the pandemic's spread, including the use of real-time location data for contact tracing.

Figure 1. Big Tech Market Capitalization
(Billions of US dollars)



Source: Bloomberg Finance, L.P.

2. The use of individual data can generate important economic and financial benefits but also pose challenges. The proliferation of data in the economy presents an opportunity to boost growth through efficiency and innovation and to increase access to financial services by reducing information costs. This is because data's economic properties are unlike those of other inputs—including labor and oil. Data is nonrival: the same data can be used by many people simultaneously without being depleted. The more it can be shared, the higher the social returns from innovation and knowledge generation and the greater the reduction of the information cost of finance. When individual data is exchanged, however, the transaction affects people's privacy and can leave them at a strategic disadvantage. Without awareness and compensation, the data market could leave some of us worse off. Moreover, use of individual data can allow for more targeted discrimination because of previously private personal characteristics.

3. **Governments must balance many objectives when setting data policies.** Public interest in the rules for the data economy is building across the IMF membership: many countries are considering or have passed new laws governing how personal data can be collected, processed, used, and shared. Modernized data policy frameworks need to address two concerns about the status quo. First, data markets are too opaque—most of us participate in the modern digital data economy daily, but we aren't fully aware of how our data is used, transferred, and processed. To achieve an efficient and equitable data economy requires clear rules and effective consent on use and processing of data. Second, companies that build large data sets have a strong incentive to hoard them. This potentially stifles competition and innovation and reduces the social benefits that could flow from wider data access. Moreover, there is a strong public policy case for promotion of the resilience of the digital data infrastructure, ensuring data integrity and protecting the data held by public and private entities from theft and misuse. Shortcomings here threaten public trust and financial stability that policy measures should mitigate, including adequate investment in cybersecurity. All these issues need to be balanced against the public interest in the disclosure and sharing of certain types of data, including for regulatory and enforcement purposes.

4. **Dealing with these challenges will require greater coordination first among domestic regulators.** Data policies impose trade-offs that affect growth, privacy, competition, and financial stability and integrity. Managing these objectives has traditionally been assigned to separate ministries and regulators. Effective data policy will require a coordinated approach that brings many actors to the table to manage complex trade-offs. Mechanisms to foster greater coordination between regulators merit further exploration; focusing on single objectives can have repercussions for others.

5. **Global cooperation is also needed to contain the fragmentation of the digital economy across national borders.** Data is a highly mobile resource, and the potential for economies of scale brings enormous potential gains from cross-border data flows. But if countries don't trust how global partners handle data, or if they feel they aren't deriving enough benefits from its exploitation, they may opt to erect digital barriers that impede international data flows, undermining innovation, financial stability, integrity, and efficiency. If the global digital economy fragments, smaller countries could be at a

disadvantage, as they would be less able to access large data pools in major economies needed to compete in providing data-intensive services.

6. This Staff Discussion Note brings together the many strands of discussion on the data economy and discusses its macro-financial implications and the case for policy cooperation. It discusses what data is (Annex I) and sets out a macro-financial framework to explore the economic and financial effects of using data. It then delves deeper into the role of data in financial services and the financial stability implications, given the importance of these issues to the IMF's mandate. The note then discusses national approaches to data policy frameworks and the need for domestic policy coordination. It concludes with a discussion of data in the global economy and the case for international cooperation.

7. The broad scope of this paper is meant to inform the IMF's digitalization work, including on digital money, taxation, and competition in the digital economy. As central banks consider the launch of digital currencies, they stand to become focal points for massive data flows, presenting novel opportunities and risks that affect policy objectives outside the central bank's mandate. The cross-border adoption of private digital currencies and digital payment solutions will be heavily influenced by national data policies (IMF 2020). Discussions of taxation of the digital economy require careful analysis of the value of data (De Mooij, Klemm, and Perry 2021). Finally, analysis of competition in the digital economy is of growing macroeconomic relevance (Akcigit and others 2021). In all these cases, remedies to specific problems arising from the proliferation of data can generate unexpected trade-offs elsewhere in the domestic and global economy.

THE ECONOMICS OF DATA

8. Data matters—but how exactly? A rapidly growing body of literature in economics and finance has studied the roles of data in the economy and the effects that may emerge when data proliferates.[2] What does data do in the economy?

[2] Carrière-Swallow and Haksar (2019) provides a comprehensive summary of the literature on the economics of data from a public policy perspective. Another recent review is provided by Veldkamp and Chung (2019).

9. Data is an input in the production of goods and services, particularly relevant in the digital economy. Deriving value from data as an input requires costly processing and analysis so that it can be used in combination with other factors of production, such as labor and algorithms. This is a salient way of thinking about the role of data used in artificial intelligence (AI) applications. In this function, data analysis is used as part of the innovation process, with new insights and predictions leading to the development of new products and services.

10. Big data has supported AI in solving increasingly complex problems. The proliferation of big data and the development of more sophisticated and flexible machine learning algorithms have enabled data analysis to address increasingly complex problems. AI is being deployed as a general-purpose technology in an increasing number of fields to tackle very diverse problems (Agrawal, Gans, and Goldfarb 2018; Boukherouaa and Shabsigh, Forthcoming). For instance, a car equipped with sensors may record the actions of a driver navigating city streets, building up a massive data set of human decisions in the face of various situations. Patterns in this data can then be analyzed using machine learning algorithms to predict and mimic human decision-making in complex road environments, which may then enable the production of a safe self-driving car.

11. Data is a *nonrival* input, in the sense that it can be used multiple times and by multiple agents simultaneously without being diminished. This characteristic gives data important implications that set it apart from other inputs, such as labor, capital, and oil, whose use is limited. One important insight is that society will get more benefits from the data it generates when it is made widely available to many data processors (Jones and Tonetti 2020). But will it be widely available? When a data processor has collected data valuable to its commercial interests, there is a strong private incentive to hoard that data and withhold it from competitors. And should data be widely available? Granting broad access to individual data may generate more value but may also compromise privacy and heighten cybersecurity risks.

12. Individual data contains information about consumers and firms, and access to it shifts that information. When access to data reduces information asymmetries between buyers and sellers, it can lead to more efficient economic transactions. For instance, a company with access to data about the

characteristics of potential consumers—such as their interests and buying habits—can use that information to offer a more personalized good or service, such as an advertisement for a product that consumers are more likely to find useful or desirable, or a lower car insurance premium in return for permission to be tracked while driving. This type of personalization can potentially make both customers and the seller better-off, if they have control over the use of their data.

A. The Value of Data and Who Should Derive It

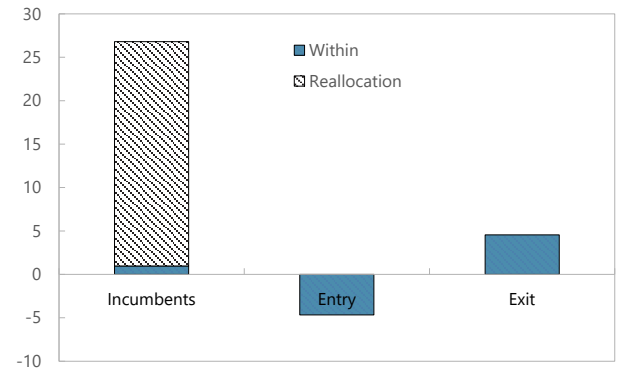
13. Assessing the value of individual data is difficult. Unlike a commodity, data is highly heterogeneous. Since no two pieces of data are perfect substitutes for each other, they need not hold the same value, and their value may change over time. Moreover, individual data transactions usually take place through barter: individuals swap use of their data for "free" digital services. As such, assessing data's value is challenging, even for those who have direct incentives to do so. For instance, advertisers spend large sums on data about online users, on the premise that displaying a well-targeted ad will influence the behavior of the person seeing it. Some studies flag the difficulty of measuring the returns to such efforts (Lewis and Rao 2015) and point out that the returns may in fact be negative (Marotta, Abhishek, and Acquisti 2019). Information from stock listings, mergers, and acquisitions offers a sense of the commercial value placed on data, but may not provide a full picture of its social value. This is because the data economy is opaque, and privacy is not fully respected. A lack of information, combined with individual costs of privacy breaches that are not fully internalized by data processors, may lead to data with a price that differs from its true economic value. Governments also remain divided over how to place value on data from a digital taxation perspective (Aslam and Shah 2021).

14. Who will accrue the returns to data? A key question is how much of the value of data comes from each individual data point and how much from its aggregation and subsequent analysis. The answer is likely to differ according to the type of data and the context, but an important factor is the degree of substitutability between data and other factors of production. For instance, some types of data may require very advanced and proprietary analytical tools to convert it into useful information. In the case of big data sets used to train machine learning algorithms, a question is how much of the value comes from the analysis provided by the highly skilled labor required to process the data. In other cases,

information may be extracted with less analysis, as when individual data is used to provide a targeted product. The value of such data may differ greatly depending on characteristics of the data subjects, such as their income, age, and consumption preferences.

15. The allocation of value in the data economy will depend on competition and the market power enjoyed by individuals and data processors. If a data processor enjoys market power, then obtaining granular information on its clients may make it possible to extract considerable rents through the implementation of price discrimination strategies, a practice that has been documented for major online retailers (Hannak and others 2014). Data may also represent a source of market power if a stockpiled data set acts as a barrier to entry that deters competition. Indeed, the tech sector has displayed rising market shares for firms with strong market power, consistent with a winner-takes-most market structure. Akcigit and others (2021) document this phenomenon in a study spanning 82 countries, showing that the main driver of rising markups in the tech industry is the rising market share of existing high-markup firms (the "reallocation effect" shown in the Figure 2). This is also relevant in finance, where banks have traditionally been able to hoard the data generated by their relationship with individuals and companies, making it hard for other lenders to accurately price the risk associated with extending a loan and thus being less able to compete with incumbent banks (Carrière-Swallow and Haksar 2021a).

Figure 2. Decomposition of Markup Increases in the Tech Industry
(Percent)



Source: Staff Discussion Note No. 2021/001 , International Monetary Fund.

B. Privacy and the Digital Economy

16. Decisions on privacy have important economic consequences. The ability to share data over digital networks and to make it public to a global audience has increased the salience of privacy (Acquisti, Taylor, and Wagman 2016). Significant benefits can accrue to individuals from the use of their

data, including innovative services and more customized products. But when these individuals are unaware, or don't have a say when it comes to the use of their data, an externality results: decisions by companies about whether to collect, process, or share personal data can harm the individual, who may not be compensated.[3] These externalities are often negative, with the data used for instance to charge the individual a higher price, and this leads to too much data on individuals being collected and processed (Acemoglu and others, forthcoming). An emerging example that will give rise to privacy issues is the shift away from cash and toward digital payments and digital money—including central bank digital currencies—whose inherent traceability calls for serious decisions about the privacy accorded to participants in transactions (Garratt and van Oordt 2021).

17. **Effective privacy is about giving individuals agency over their data.** Privacy should not be understood as preventing the sharing of personal information, but rather as giving data subjects control over access to their data (Acquisti, Taylor, and Wagman 2016). When companies are given the right to govern data access, the result may be less competition, more data hoarding, and less privacy for consumers (Jones and Tonetti 2020). For the data market to work more efficiently, data subjects must be able to adequately control access to their data to close the externality discussed above. Nonetheless, there can be a conflict between an individual's desire to restrict data sharing and the public good, such as when individuals refuse to participate in contact tracing programs in the context of a pandemic.

18. **Placing a value on privacy is inherently difficult.** Research in the literature on privacy has identified an apparent privacy paradox (Nissenbaum 2009): people place a much lower value on their privacy when it comes to their private actions than they do when asked in a survey to give it a subjective value.[4] While a large percentage of people tell survey takers they are very concerned about a company

[3] Harm from sharing of data is caused by undermining data subjects' preference for keeping their personal characteristics or actions private. In addition, the data may be used strategically by those that acquire it to extract rents from the data subject, including through price discrimination.

[4] In discussing survey evidence on consumers' stated valuations of privacy, Winegar and Sunstein (2019) argue that information deficits and behavioral biases render these valuations uninformative about the true economic value of privacy.

sharing their private information, almost all willingly give their consent to do so in exchange for the most basic of "free" online services.[5]

19. Anonymization can enable some of the social benefits of data access while preserving privacy.[6] In several applications, data analytics can provide valuable insights without data being individually identifiable. Consider training an AI system to drive an automated car or studying the effects of new vaccines against pandemic diseases based on anonymized data from around the world (Box A1). These applications rely on huge amounts of individual data to develop AI algorithms, but do not require that the data be linked to an identified individual. But in many other applications, the value of data is substantially reduced through anonymization, precisely because it no longer reveals information about a specific person.

DATA AND THE FINANCIAL SECTOR

20. Access to personal data is a core input in the provision of financial services. Lenders always face uncertainty about the creditworthiness of their clients. The provision of financial services has always relied on data to reduce these information gaps. This need for data has intensified in recent decades as the global financial system has become heavily digitalized and interconnected. The collection of data for regulatory purposes also has been steadily growing as financial services providers are subject to increasingly stringent regulatory requirements to identify customers, detect suspicious transactions, and ascertain the source of funds.

[5] Johnson, Shriver, and Du (2020) find that only a tiny share of American users—representing 0.2 percent of ad impressions—opted out of targeted advertising under the industry self-regulated AdChoices program. Aridor, Che, and Salz (2020) study the implementation of the General Data Protection Regulation (GDPR) in the European Union, which mandated that websites offer their users the ability to opt out of being tracked, and find that only 12½ percent of users of a particular website opted out even though it meant that the ads displayed were less precise. This suggests that a significant share of users is privacy conscious, but also that the majority of users remain willing to engage in a standard data-for-service transaction even when offered the chance to opt out. A salient example is mobile weather forecast applications; users eagerly grant access to a stream of detailed location data just to avoid typing in the name of their current city (Carrière-Swallow and Haksar 2021b).

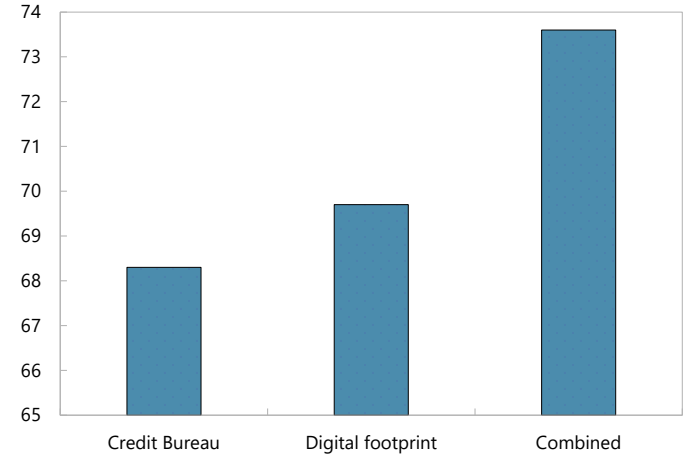
[6] There could however be limits to anonymization as it may be possible to match other characteristics of individuals to identities to effectively reverse anonymization (Su and others 2017).

A. Data and Efficiency in Financial Services

21. Incomplete information about borrowers prevents the efficient allocation of credit because of adverse selection. Suppose an unknown customer walks into a bank—will that person be offered a loan on good terms? Stiglitz and Weiss (1981) show that, when lenders do not have full information about borrowers' creditworthiness, they are likely to ration credit—that is, some borrowers will not be offered a loan at any interest rate and will thus be excluded from the financial market. This is because borrowers' willingness to accept a very high interest rate on a loan signals to the lender that they are unlikely to repay it: the higher the interest rate offered by the lender, the riskier the pool of borrowers willing to accept it. Adverse selection presents a particularly relevant friction in developing economies, where potential borrowers working in the informal sector or without previous access to financial services are often turned away by banks that cannot evaluate their ability to repay.

22. Access to nontraditional data can boost inclusion by alleviating adverse selection problems that exclude disadvantaged populations from credit markets. As discussed in the Bali Fintech Agenda (IMF 2018b) and Sahay and others (2020), the availability of more data about borrowers is a key element of the promise that technology offers to the provision of financial services. Information collected in the context of online services, including social habits, payment of utility bills, and other traces of economic and social activity, may form the basis for evaluating the creditworthiness of a borrower who has not had previous interactions with a financial services provider. For instance, Frost and others (2019) and Berg and others (2020) present evidence that

Figure 3. Explained Credit Scores (area under curve) with Different Models



Source: Berg, T., Burg, V., Gombovic, A., and Puri, M. (2020).

nontraditional data collected online can predict creditworthiness more accurately than a traditional credit score (Figure 3).[7]

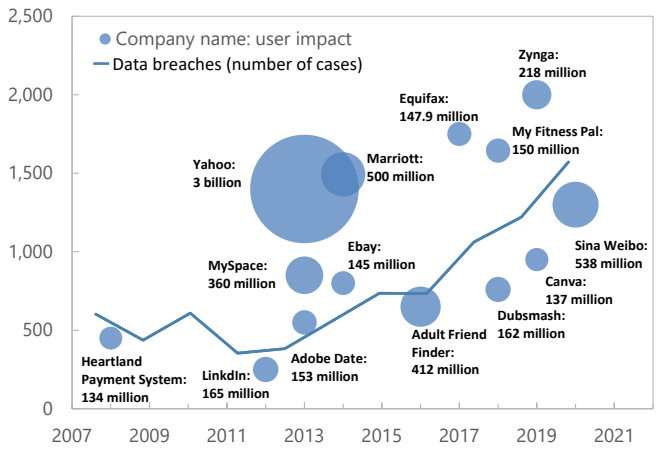
23. New approaches to data sharing give individuals greater agency over their data and can boost competition and efficiency. Data sharing in financial services has a long history. Since the end of the 1800s, credit bureaus have operated as data brokers with which lenders share information about borrowers on a reciprocal basis—that is, a bank that provides information about its borrowers is entitled to receive certain classes of information about a prospective client. Where private credit bureaus have not emerged spontaneously, many jurisdictions operate public credit registers that mandate the sharing of default information about borrowers. Data sharing among financial services providers can lessen adverse selection and moral hazard problems and thus increase the provision of credit by reducing information asymmetries. Jappelli and Pagano (2002) and Djankov, McLiesh, and Schleifer (2007) present empirical evidence that data-sharing institutions generate deeper and broader credit markets and less frequent defaults. This desire for greater competition and efficiency in financial services is behind the emergence of open banking initiatives around the world (Box A2).

[7] Figure 3 draws on Berg and others (2020) to illustrate the gains from explaining the scoring of individual credit using a basic credit bureau scoring model contrasted with using publicly available basic individual data from people's "digital footprint" (including the type of phone owned, and hours of online shopping) and last by combining both approaches. The use of big individual data significantly improves the ability to judge credit quality. Use of such alternate data could also attenuate the concerns raised by some that limiting credit analysis by intermediaries to traditional variables in the run-up to the global financial crisis undermined risk assessment.

C. Data Resilience in Financial Services

24. Data resilience is needed to ensure financial stability. The financial system relies on a robust and resilient digital data infrastructure to perform its key economic functions. Every day this infrastructure facilitates financial transactions between market participants and end users, requiring data to flow seamlessly through the network. Interconnected and interdependent information and communications technology (ICT) systems need to process data reliably so that the financial system can perform its key economic functions. Moreover, these systems rely on sensitive individual data, which need to be protected. These needs are illustrated by the sharply rising incidence of data breaches over time, with some very large exposure of individual data records (Figure 4).

Figure 4. Cybersecurity and Data Breaches



Source: Hill and Swinhoe, 2021; and IMF staff calculations.

25. Private incentives to protect data may not deliver systemic data resilience. Confidentiality (data access is not granted to unauthorized individuals or systems), integrity (that which is recorded and shared is indeed true, accurate, and complete), and availability (the property of being accessible and usable on demand) are three crucial properties for the data processed by financial systems. The financial system depends on public trust, which can be threatened by cyber incidents. Private incentives for firms to invest in cybersecurity may not consider the impact of a data breach on public trust in the broader system, leading to underinvestment and systemic cyber risk (Kopp, Kaffenberger, and Wilson 2017; Kashyap and Wetherilt 2019).

26. The growing use of cloud services to store and process data introduces new trade-offs. Third-party cloud service providers are growing as critical repositories for data held by financial intermediaries, offering more efficient data management, analytic tools, and cutting-edge cyber defenses.

But the attendant massive agglomeration of data in a handful of cloud service providers poses potential systemic financial stability risks from a single point of failure. The Financial Stability Board (FSB) (2017, 2019) has noted that use of such services may reduce operational risk at the individual firm level—for example, by increasing cyber resilience and supporting business continuity—but could "also pose new risks and challenges for the financial system as a whole." [8]

27. Data-intensive provision of financial services could increase concentration and impact financial stability. Large intermediaries may generate more data and may gain advantage over smaller competitors. The resulting tendency toward greater concentration in financial services, beyond a certain point,[9] could pose stability risks. Indeed, concentration of data in just a few global large platforms could lead to an unprecedented degree of concentration—or "too-big-to-fail"—risks. Moreover, the efficiency gains from improved data proliferation in finance may be unevenly shared across borrowing firms, benefiting larger incumbent firms more than smaller firms, which produce less data as a by-product of their operations (Begenau, Farboodi, and Veldkamp 2018). A rich data vapor trail thus allows larger firms to lower their cost of finance, reducing their cost to finance expansion, which in turns generates more data. The implication is that data-based lending will tend to favor more concentration in production (Farboodi and Veldkamp 2020). In addition, the recent expansion of big-data-driven credit models and the fact that the data used do not, in many cases, span a full financial cycle, raise questions about the performance of this new lending if economic and financial conditions were to deteriorate (Claessens and others 2018).[10]

[8] Beyond the need for coordination domestically and internationally, a further challenge for financial authorities is to deepen the scarce skills needed for effective supervision of technology providers, which lie at the intersection of economics, finance, and computer science.

[9] Vives (2016) discussed a view on the hump-shaped link between competition and stability in the financial sector, along the lines that some concentration, up to a point, is good for stability; beyond that point, more concentration is bad for stability.

[10] It is worth noting that new research examining the experience of China during the COVID-19 crisis indicates that digital banks' portfolio quality has held up very well (Sun and others, 2021). Moreover, recent IMF research (Pierri and Timmer 2020) finds that banks that adopted data technologies more intensively before the global financial crisis were more efficient in screening borrowers and thus experienced fewer mortgage defaults and problem loans during that crisis.

28. Data is also increasingly gathered in the financial sector to satisfy regulatory requirements. Financial services providers, including banks and other intermediaries, have an important responsibility to safeguard the integrity of the financial system; in a country's regulatory framework, this responsibility takes the form of obligations to apply preventive measures aimed at insulating the financial sector from criminal activity and illicit flows. Due diligence procedures, such as customer identification and transaction and customer monitoring, by their nature rely heavily on data. The volume of data to be processed to meet legal requirements means that these measures increasingly require high-quality IT systems both for market participants (for example, AI to track spending patterns to detect outliers and suspicious transactions) and for regulatory authorities (for instance, "suptech," the use of technology to support financial supervision).

D. Data and Discrimination in Financial Services

29. The use of granular individual data could lead to the exclusion of people who exhibit traits associated with risky financial behavior. Consider a health insurance company that can harvest customer data to build a profile of preexisting medical conditions or risk propensities that are not necessarily linked to risky or harmful behavior such as fast driving or heavy consumption of sugar-rich foods. This data analysis of underlying health conditions may lead an insurer to charge higher premiums or deny coverage (Arrow 1963), potentially excluding vulnerable individuals from insurance markets. It is conceivable that the unfettered availability of very granular data could undermine the risk-sharing function of insurance in many instances. To mitigate such an outcome, regulations should specify the types of data that may be used to make decisions—for example, restricting in some cases the denial of health insurance coverage based on a preexisting condition.

30. Inappropriate use of individual data to train AI models could worsen biases in decisions on access to financial services. AI algorithms that increasingly power lending, investment, and wealth management decisions are trained on the data of "expert individuals." Absent appropriate governance of the data that is used to train these algorithms, bias inherent in the training data set could be injected into algorithms. Consider a pattern of lending that effectively reduces access to individuals based on location or race. The risk that the use of AI in lending may worsen racial biases in the allocation of credit is vividly

demonstrated in a study by Fuster and others (2020) of the US housing market. Furthermore, the use of AI algorithms—which produce accurate predictions but often lack a structural interpretation—may be perceived as a discriminatory black box that loan officers will not be able to explain to their customers or to regulators.[11]

THE DATA POLICY TOOLKIT

31. Data policy frameworks set the rules governing the use of individual data in the economy. When well designed, these policies can underpin public trust and participation in the digital economy (World Bank 2021). These frameworks include the design of standards and policies, as well as their enforcement, on how data is stored, who may have access to it, how it can be used, and for what purposes it may be used (Box A3). With the proliferation of digital services, data governance has become increasingly important to the formulation of public policies across multiple objectives. Countries' approaches to data governance have important international ramifications because cross-border provision of data services must comply with local as well as foreign frameworks, generating spillovers from data policies. Individual jurisdictions tend to prioritize domestic considerations and not the impact of their policies on other countries.

32. National approaches differ widely, with three broad trends apparent in the global digital economy.[12] Addressing privacy concerns and protecting the data of nationals have been important drivers of most data regulatory frameworks. A large part of data regulation stems from legal—and in some cases, constitutional—concerns regarding privacy. Underlying this is the debate over the extent to which individuals should have agency over the use of their own data. Rights-based approaches, like the European Union's approach in the General Data Protection Regulation of 2016, stand in contrast to the

[11] Such biases are not restricted to finance. Buolamwini and Gebru (2018) find that facial recognition error rates of major tech companies' systems in the identification of darker-skinned people were dozens of percentage points higher than for lighter-skinned people. The issues lie in part with the data sets used to train the systems, which can be overwhelmingly male and white.

[12] Differences in national frameworks can result in regulatory arbitrage or forum shopping (seeking the jurisdiction with the most attractive legal framework or jurisprudence). To manage regulatory gaps in cross-border data transactions, parties can contract—or litigate. From a private law perspective, the foundational considerations in a case are the appropriate jurisdiction and applicable law. So far these are mainly based on the location of the individual whose data privacy has been impacted, rather than the location of the company.

more activity-based regulation in the United States. The US open transfer approach is largely in the spirit of industry self-regulation, based on the concept of "notice and choice" (World Bank 2021), and data privacy protections in the United States are typically sector-specific and apply to a relatively narrow field, such as health care or finance. Some large emerging markets—China, for example—protect privacy for private individuals but also emphasize a public interest in being able to maintain access to data on individuals and have introduced data localization to protect the data of individuals. In part because of the weight given to these different factors, there is no overall clear approach to a global data governance framework, with attendant risks of fragmentation.

33. Distinguishing between public and private law is important for understanding the legal underpinnings of data frameworks.[13] A significant portion of private law applies to enforcing contractual agreements between private parties on data usage, typically involving certain forms of consent. Public law is regulatory in nature and aims to set standards and rules for data usage. Private enforcement (for example, allowing redress of contract violation via civil lawsuits) is often focused on whether data was used in the manner for which consent was provided. Public law provides fundamental rights that are considered inalienable and cannot be contracted away (for example, privacy under the GDPR). The degree to which consent is required for certain types of usage once an individual or entity "owns" data has also generated controversy linked to the debate over whether a property right over data is possible (see Box A4).

34. Privacy considerations must be balanced against public interests, which often necessitate the sharing of information, between private actors as well as with officials. Situations exist where an individual's private or confidential data has social value (for example, for identification purposes), particularly in sectors that have been demonstrated to be at risk of misuse (such as in the financial services sector). The social value obtained from disclosure of such information (mitigation of financial stability and integrity risks) must be weighed against the right to privacy of the individual. This balance is generally woven into public law frameworks in the form of disclosure requirements as a prerequisite for

[13] Broadly speaking, private law focuses on the horizontal claims and liabilities arising from the relationship between particular individuals and entities (and seeks to address "private" harms); public law focuses on the vertical rights and obligations of individuals in their relationship with society at large and with the state (and seeks to benefit the public good or correct "public" harms).

the acquisition or access to certain services (for example, financial services) or fulfillment of certain legal obligations (such as tax filing or suspicious transaction reports). Disclosure may also be required upon demand (by regulatory or law enforcement authorities).

35. There is a strong case for greater regulatory coordination. Sectoral data policies create economic trade-offs that could be managed better. Tighter privacy protection can limit access to data, stifling innovation.[14] Boosting competition in finance through open banking could create some risks for financial stability. An approach in which a sectoral regulator considers privacy, innovation, competition, or financial stability objectives, in isolation, may well have unintended consequences for other objectives. Balancing these competing objectives calls for a coordinated approach to data policy frameworks that involves cooperation among many national agencies, including central banks, ministries of finance and economics, financial regulators, consumer protection agencies, privacy regulators, and competition agencies. This would require, at minimum, consistency across relevant legal and regulatory frameworks as well as some form of institutional arrangement for discussion and consultation among regulators. For example, in the context of issuing digital monies to further their monetary and financial stability objectives, central banks will need to consider how to handle privacy issues arising from the gathering of individual payment information, either by themselves or by private entities handling the digital monies (Bank of England 2021). This could require a mechanism for coordination with domestic privacy and consumer protection regulators. However, balancing objectives becomes even more complex in a cross-border context, with each country having its own policy objectives and data policy frameworks.

36. Efforts to modernize data policy frameworks should start by clarifying the rules of the digital economy, while ensuring that it is competitive and resilient. For the market to function efficiently and for value from data to be properly measured, participants in the digital economy should know and understand how their data is being collected, processed, and accessed. Continued innovation

[14] For example, observers have flagged a concern that the GDPR's emphasis on privacy protection may reduce innovation in Europe by acting as a tax on web-based technologies or that the compliance costs for start-ups may be very high, reducing competition and thus hurting consumers. There is early evidence that the GDPR has had a small but significant effect on the ability of e-commerce firms to generate revenue from users (Goldberg, Johnson, and Shriver 2020) and to raise funding (Jia, Jin, and Wagman 2021), and that it has increased market concentration among web technology vendors (Johnson, Shriver, and Goldberg 2020). During the pandemic, differing privacy protection standards have made it harder to collaborate on crucial medical research across borders because of the difficulty of sharing individual results of biomedical trials (Peloquin and others 2020).

and greater equity also require a competitive digital economy, and concerns have been raised about the concentration and high markups observed in tech-heavy sectors (Furman and others 2019; Akcigit and others 2021). Several policy approaches have been proposed:

- **Interoperability:** Interoperable platforms provide a mechanism for ease of sharing and transmission of data pursuant to a mutually accepted set of rules. These platforms harvest the data of a large network of individuals. Indeed, separating the value of the network from that of the underlying data in the platform is hard. But the two are closely linked. Thus, tools that make it easier for individuals to move between networks (so called multi-homing) go a long way toward reducing the hold of networks and increasing contestability.[15] This could be pro-competitive and also supportive of stability in the financial sector were it to lower "too-big-to-fail" risks. To achieve interoperability, platforms must be accessible by relevant actors; rights and responsibilities also must be assigned and managed.
- **Data portability:** Although interoperability can be thought of as a network of pipes that connect different platforms, portability gives people control over what can flow through those pipes. An aim of portability requirements—as conceived for example in open banking arrangements—is to promote competition by making it less costly for a data subject to switch to a competing service or to multi-home across multiple services. There is scope to consider portability as a solution to managing competition more broadly in network environments.
- **Data fiduciaries:** The concept of data fiduciaries—agents that have the responsibility to manage the subject's data and rights and seek consent for data processing—has been advanced as a potential solution to the problem of effective consent management, including proposals discussed in India.[16] These could be supported as solutions to achieving privacy objectives while operationalizing the benefits of wider consensual data sharing.

[15] The 2020 proposals on the Digital Markets Act and the Digital Services Act introduce several of these proposals, including on third-party interoperability requirements for Big Tech gatekeepers—including social media and online marketplaces—in certain situations and enhanced consent management and protections. But these are again largely framed with domestic policy objectives in mind.

[16] See Carrière-Swallow, Haksar, and Patnam (2021) for a discussion of the approach considered in the case of India. Acemoglu and others (forthcoming) show that data intermediaries could mitigate the externalities that emerge in unregulated data markets by allowing for authorized and desired data sharing without revealing information about other individuals.

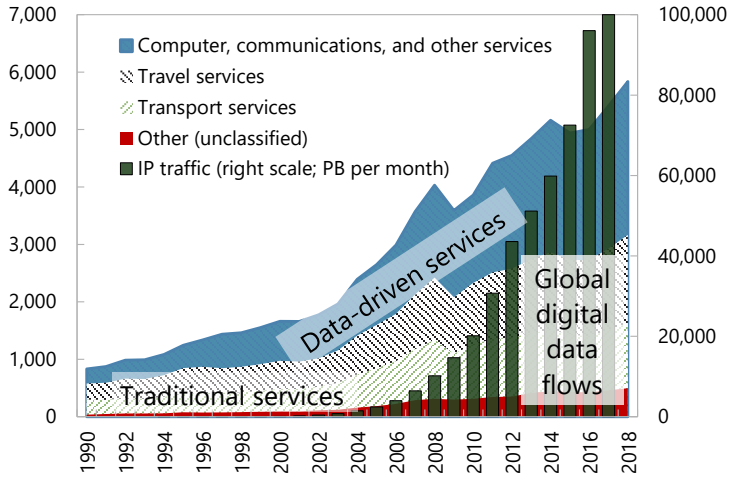
- **Public data utilities:** Greater data sharing can be in the public interest, including in the context of pandemic contact tracing and biomedical research. There can be scope for multistakeholder solutions to data sharing while preserving privacy, such as the Pan-European Privacy-Preserving Proximity Tracing (PEPP-PT) community initiative.[17] Aggregation of individual data for legitimate public good in an independent central repository (along the lines of a public credit bureau model) is worth considering. This could allow level-playing-field access to data for the development of solutions for businesses of all sizes, but it would have to address surveillance state and cybersecurity concerns and may be dominated by decentralized solutions such as interoperability and portability.

[17] PEPP-PT seeks to develop open protocols for contact tracing that are GDPR compliant and allow individual data to be stored only on mobile phones, yet facilitate anonymized tracking using Bluetooth proximity sensing.

THE CASE FOR GLOBAL POLICY COOPERATION

37. Data has become the ultimate modern mobile factor, and cross-border data flows are rising. The ability to move data across borders underpins a growing range of economic activity and international trade (Figure 5; data on trade in computer, communications, and other services versus traditional services trade). Data flows are particularly crucial for trade in services and to facilitate cross-border payments, whose high costs reflect in part disparate standards on the format for exchanging data on identity and payment details.[18]

Figure 5. Global Trade in Data-Driven Services
(Billions of US dollars, unless noted otherwise)



Source: World Integrated Trade Solutions (WITS); and World Bank (2021).
Note: IP = internet protocol; pb = petabyte.

38. Cross-border data protection presents challenges. One challenge is that citizens' personal data can flow in and out of jurisdictions that do not offer comparable levels of privacy protection.[19] It could also be argued that with incomplete markets, national data subjects are not being fully compensated for the processing and use of their data by globally based corporations. Authorities also have a legitimate interest in maintaining control over individual data that may be needed for regulatory or security purposes, generating tensions between privacy and other policy objectives that may be

[18] The global community is considering how to reduce these costs, including using digital currencies (IMF 2020) and through improvements in digital payments infrastructure (Committee on Payments and Market Infrastructures 2020). Harmonization of data standards could enhance cross-border payments and have widespread benefits. Promoting the adoption of common data formats, as well as protocols for information exchange, can reduce costs and improve the scope for straight-through data processing in existing payment systems and arrangements.

[19] Under the GDPR, personal data may be transferred out of the European Union only if the other country provides comparable data protection. While this is a particular challenge for developing economies (Mattoo and Meltzer 2018), it has also led to tensions with major advanced economies. A recent example is the Schrems I and Schrems II litigation following the disclosure that data of non-US Facebook users could be accessed by US intelligence agencies, which led to the breakdown of the EU-US Privacy Shield agreement.

perceived differently across countries. Garcia-Macia and Goyal (2020, 2021) point to conditions that may lead countries to seek to erect barriers against trade in technologies in monopolistic sectors, such as in the digital economy, to ward off challengers. They warn that these decisions could lead to the digital decoupling of major technology centers from each other and from the rest of the world.

39. Differences in data standards pose particularly stark trade-offs for developing economies. Cross-border data flows drive the most dynamic exports of developing economies: data processing and data-related business services. These services, ranging from financial accounting and tax returns to medical transcription and diagnostics, contributed to more than $50 billion worth of developing economy exports to the European Union in 2015—of which one-fifth came from Africa (Mattoo and Meltzer 2018). Developing economies thus face a dilemma when advanced economies tighten data regulation: either they must adopt these tighter standards—with higher compliance costs for their exporters—or face losing market access.

40. Data localization laws could generate broad economic costs and harm smaller economies disproportionately. Several countries are seeking to place specific limits on the transfer of data on their national subjects outside their national boundaries (so called data localization). Menon (2018) points out that data localization policies may reflect misguided concern about the cybersecurity risks of cross-border data or flows or protectionist policies and could undermine the benefits of digital trade. Since data is nonrival, there are potentially large gains from its cross-border usage. Economies of scale may also emerge as data sets grow large enough to tackle frontier AI prediction problems. Reducing trade in large data sets could thus undermine economic growth. Firms in smaller countries facing strict data localization requirements in their trading partners could also find it increasingly difficult to compete and innovate without access to large data sets, which could generate a digital divide that excludes some countries from the benefits of digitalization.

A. Toward Global Data Policy Frameworks

41. There is a strong case for international cooperation on data governance. While we should not expect all countries to handle issues of innovation, privacy, and security the same way, international dialogue and cooperation can ensure that the digital economy does not become subject to undue fragmentation. Aspiring to the best principles of privacy and individual rights, while satisfying social objectives, need not set off a global scramble to fragmented policy approaches leading to localized data markets that could undermine the many potential benefits of cross-border data sharing. There is a need for common minimum principles across countries that balance the interests of growth and competition with national and individual privacy concerns, and it is paramount that this be done in a setting where all countries can have a say, to avoid the emergence of a digital divide (World Bank 2021).[20]

42. Key elements of common minimum international principles—especially for financial services that are already highly regulated with significant international coordination—could include the following:

- **Principles for data protection:** An international agreement on common minimum standards for acceptable protection of individual data when it is shared across borders would reduce uncertainty for businesses seeking to comply. Such an approach could draw on the OECD Principles on Privacy (1980 and amended 2013), with further thought on issues such as the role of consent and the definition of data and individuals.
- **Principles on interoperability and data portability:** Given the global reach of businesses that make use of individual data as an important part of their business, there is a need to discuss common principles on how such interoperability and portability should work across borders (Furman and others 2019). Specific use cases include cross-border payments and cross-border data sharing across open banking initiatives (Committee on Payments and Market Infrastructures 2020). A concrete challenge is to coordinate on principles for enabling the interoperability of digital currencies issued by central banks when these can be used across borders, including a method for digitally identifying individuals and standards for digital wallets and data flows.

[20] See for example proposals for appropriate use of customer data in the financial sector put forward by the World Economic Forum (2018) focusing on consent, control, security, transparency, and reciprocity.

- **Principles on data sharing for regulatory purposes:** A rigorous data framework should govern not only the protection of data but also its disclosure to public bodies, including regulatory authorities, where necessary to meet certain public policy objectives (for example, to facilitate criminal law enforcement activities and determine tax liability). Exemptions to confidentiality and secrecy provisions are already commonplace in many national frameworks (for instance, tax law and anti–money laundering and combating the financing of terrorism) and in line with many international standards and best practices (such as Financial Action Task Force standards for combating money laundering and the financing of terrorism and Organisation for Economic Co-operation and Development (OECD) tax initiatives). Carefully balancing privacy concerns against disclosure for public policy objectives continues to be needed as data regimes evolve. In line with current efforts, and to the extent possible, principles on data disclosure to public authorities should aim to achieve common ground across national frameworks to allow for global data sharing for enforcement or regulatory purposes.

43. The G20 has also recognized that there is a need for access to private data sources for public policy purposes. Policymakers and statisticians continue to face barriers to data access from private entities, domestically—because of issues such as confidentiality or legal provisions—that become even more complex across borders. In April 2021, the G20 finance ministers and central bank governors asked the IMF, in close cooperation with other international organizations, to prepare a proposal for a new G20 Data Gaps Initiative, in which *Access to Private Data Sources* is featured as one of the four main priority areas.

44. A few global and regional data frameworks have already been discussed. Some limited formal arrangements for exchange of data for the provision of commercial and financial services are governed by multilateral trade treaties, particularly via the World Trade Organization (WTO), and bilateral trade agreements. While these can be binding, most rely on participants to opt in, and some major countries have elected not to do so. There are also bilateral agreements and statements of intent that are not binding but are effectively hard law (for example, the EU-US Privacy Shield agreement, though it was invalidated following the Schrems litigation). Certain global frameworks are also in place for the large-

scale exchange of data between countries for official purposes.[21] A more "soft law" approach, composed of international standards and practices, has also been proposed, especially with regard to data privacy (for example, the previously cited OECD guidelines on privacy and the APEC Internet and Digital Economy Roadmap). The WTO's e-commerce initiative (a voluntary multilateral approach), at an early stage, seeks a common system of rules to allow for cross-border data flows while ensuring privacy protection.

45. Without a global approach to data policy frameworks, a continuation of ad hoc, sectoral, regional, or bilateral approaches is likely. While these raise concerns of fragmentation, a gradual and piece-by-piece approach could proceed alongside the development of common global principles. Examples include bilateral agreements that offers a modality for resolving the conflict between regulatory heterogeneity and international data flows.[22] While a new treaty regime on global data regulation seems highly ambitious**,** there is also scope to expand upon existing cooperative regimes. One approach might include building upon existing standards, such as the voluntary standards set forth by the International Organization for Standardization (ISO).[23]

46. There is a need to strike a balance at the national and international levels. Overall, this Staff Discussion Note makes clear what data policies must overcome to achieve a range of objectives. Balancing trade-offs across these objectives requires a coordinated approach at the national level, involving cooperation among different sectoral regulators. The note also presents a range of options for

[21] For example, the automatic exchange of information for "bulk" taxpayer information. There are also a number of data-sharing regimes specifically to facilitate cooperation among regulatory and law enforcement officials. Some of these are anonymized and involve primarily statistical data (for example, IMF, WTO, Bank for International Settlements); others—particularly in the law enforcement and intelligence areas—provide a forum for the sharing of personal, confidential data, including information on investigations (for example, the INTERPOL Information System, Egmont Secure Web).

[22] Whereas traditional trade agreements focused on an exchange of market access commitments, the Privacy Shield reflected an innovative bargain: the data destination country promised to protect the privacy of foreign citizens in a way consistent with their national standards; in return the source country committed not to restrict the flow of data. Such approaches have been taken in the rules on digital trade in the Comprehensive and Progressive Agreement for Trans-Pacific Partnership (CPTPP) and the United States-Mexico-Canada Agreement (USMCA) in a multicountry context. However, the privacy-security tension—and Court of Justice of the European Union findings—would remain an issue in such cases (Meltzer 2020).

[23] The ISO can and has forged industry-wide consensus when a new standard is being developed, including standards on electronic data exchange between financial institutions, data harmonization, and more. However, the ISO is not itself a regulatory authority, so using ISO standards to streamline data usage hinges on agreement by national legislators.

discussion on how to move forward to strengthen global frameworks for data sharing while recognizing national prerogatives, including on balancing public policy needs and privacy considerations. Action is needed at the national and international levels to mitigate the risk of fragmentation into localized national data pools, which would diminish the benefits that data sharing offers for productivity gains, trade, and financial inclusion. Absent new international agreements, ad hoc solutions will need to be found. In the interim, the tensions between data privacy, security, competition, and stability will continue to play out in the increasingly integrated global digital economy.

REFERENCES

Acemoglu, Daron, Ali Makhdoumi, Azarakhsh Malekian, and Asu Ozdaglar. Forthcoming. "Too Much Data: Prices and Inefficiencies in Data Markets." *American Economic Journal: Microeconomics*.

Acquisti, Alessandro, Curtis Taylor, and Liad Wagman. 2016. "The Economics of Privacy." *Journal of Economic Literature* 54 (2): 442–92.

Agrawal, Ajay, Joshua Gans, and Avi Goldfarb. 2018. *Prediction Machines: The Simple Economics of Artificial Intelligence*. Cambridge, MA: Harvard Business Review Press.

Akcigit, Ufuk, Wenjie Chen, Federico J. Díez, Romain Duval, Philipp Engler, Jiayue Fan, Chiara Maggi, Marina Mendes Tavares, Daniel Schwarz, Ippei Shibata, and Carolina Villegas-Sánchez. 2021. "Rising Corporate Market Power: Emerging Policy Issues." IMF Staff Discussion Note 21/01, International Monetary Fund, Washington, DC.

Aridor, Guy, Yeon-Koo Che, and Tobias Salz. 2020. "The Economic Consequences of Data Privacy Regulation: Empirical Evidence from GDPR." NBER Working Paper 26900, National Bureau of Economic Research, Cambridge, MA.

Arrow, Kenneth. 1963. "Uncertainty and the Welfare Economics of Medical Care," *American Economic Review* 53 (5): 941–73.

Aslam, Aquib, and Alpa Shah. 2021. "Taxing the Digital Economy." Chapter 10 in *Corporate Income Taxes under Pressure: Why Reform Is Needed and How It Could Be Designed,* edited by Ruud De Mooij, Alexander Klemm, and Victoria Perry. Washington, DC: International Monetary Fund.

Bank of England. 2021. "New Forms of Digital Money." Discussion Paper, London.

Begenau, Juliane, Maryam Farboodi, and Laura Veldkamp. 2018. "Big Data in Finance and the Growth of Large Firms." *Journal of Monetary Economics* 97: 71–87.

Berg, Tobias, Velentin Burg, Ana Gombović, and Manju Puri. 2020. "On the Rise of FinTechs: Credit Scoring Using Digital Footprints." *Review of Financial Studies* 33.

Biega, Asia J., Peter Potash, Hal Daumé III, Fernando Diaz, and Michèle Finck. 2020. "Operationalizing the Legal Principle of Data Minimization for Personalization." *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval.* New York: Association for Computing Machinery.

Boukherouaa, El Bachir, and Ghiath Shabsigh. Forthcoming. "Powering the Digital Economy: Opportunities and Risks of Artificial Intelligence in Finance." IMF Departmental Paper. Washington, DC: International Monetary Fund.

Buolamwini, Joy, and Timnit Gebru. 2018. "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification." Proceedings of Machine Learning Research 81: 1–15, 2018 Conference on Fairness, Accountability, and Transparency, New York, February 23–24.

Carrière-Swallow, Yan, and Vikram Haksar. 2019. "The Economics and Implications of Data: An Integrated Perspective." IMF Departmental Paper 19/16, International Monetary Fund, Washington, DC.

Carrière-Swallow, Yan and Vikram Haksar, 2021a. "Open Banking and the Economics of Data," Chapter 7 in *Open Banking*, Linda Jeng (ed.). Oxford University Press.

Carrière-Swallow, Yan, and Vikram Haksar. 2021b. "Let's Build a Better Data Economy." *Finance & Development* 58 (1): 10–13.

Carrière-Swallow, Yan, Vikram Haksar, and Manasa Patnam. 2021. "India's Approach to Open Banking: Some Implications for Financial Inclusion." IMF Working Paper 21/52, International Monetary Fund, Washington, DC.

Claessens, Stijn, Jon Frost, Grant Turner, and Feng Zhu. 2018. "Fintech Credit Markets around the World: Size, Drivers and Policy Issues." *BIS Quarterly Review* (September): 29–49.

Cohen, Morris. 1927. “Property and Sovereignty.” *Cornell Law Quarterly* 13 (1): 8–30.

Committee on Payments and Market Infrastructures. 2020. “Enhancing Cross-Border Payments: Building Blocks of a Global Roadmap.” Stage 2 Report to the G20, Bank for International Settlements, Basel.

De Mooij, Ruud, Alexander Klemm, and Victoria Perry, eds. 2021. *Corporate Income Taxes under Pressure: Why Reform Is Needed and How It Could Be Designed*. Washington, DC: International Monetary Fund.

Djankov, Simeon, Caralee McLiesh, and Andrei Schleifer. 2007. “Private Credit in 129 Countries.” *Journal of Financial Economics* 84 (2): 299–329.

Farboodi, Maryam, and Laura Veldkamp. 2020. “Long-Run Growth of Financial Data Technology.” *American Economic Review* 110 (8): 2485–523.

Financial Stability Board (FSB), 2017. “Financial Stability Implications from Fintech,” Fintech Issues Group. Basel, Switzerland.

Financial Stability Board (FSB). 2019. “Fintech and Market Structure in Financial Services: Market Developments and Potential Financial Stability Implications.” Financial Innovation Network, Basel.

Frost, Jon, Leonardo Gambacorta, Yi Huang, Hyun Song Shin, and Pablo Zbinden. 2019. “BigTech and the Changing Structure of Financial Intermediation.” *Economic Policy* 34 (100): 761–99.

Furman, Jason, Diana Coyle, Amelia Fletcher, Derek McAuley, and Philip Marsden. 2019. “Unlocking digital competition: Report of the Digital Competition Expert Panel.” HM Treasury, London.

Fuster, Andreas, Paul Goldsmith-Pinkham, Tarun Ramadorai, and Ansgar Walther. 2020. “Predictably Unequal? The Effects of Machine Learning on Credit Markets.” Forthcoming in the *Journal of Finance*.

Garcia-Macia, Daniel, and Rishi Goyal. 2020. "Technological and Economic Decoupling in the Cyber Era." IMF Working Paper 20/257, International Monetary Fund, Washington, DC.

Garcia-Macia, Daniel, and Rishi Goyal. 2021. "Decoupling in the Digital Era." *Finance & Development* 58 (1): 21–23.

Garratt, Rodney J., and Maarten R. C. van Oordt. 2021. "Privacy as a Public Good: A Case for Electronic Cash." *Journal of Political Economy* 129 (7): 2157–80.

Goldberg, Samuel, Garrett Johnson, and Scott Shriver, 2020. "Regulating Privacy Online: An Economic Evaluation of the GDPR." SSRN Working Paper 3421731.

Hammer, Cornelia L., Diane C. Kostroch, and Gabriel Quirós-Romero. 2017. "Big Data: Potential, Challenges, and Statistical Implications." IMF Staff Discussion Note 17/06, International Monetary Fund, Washington, DC.

Hannak, Aniko, Gary Soeller, David Lazer, Alan Mislove, and Christo Wilson. 2014. "Measuring Price Discrimination and Steering on E-commerce Web Sites." *Proceedings of the 14th ACM/USENIX Internet Measurement Conference.* ACM Digital Library.

Hill, Michael, and Dan Swinhoe, "The 15 biggest data breaches of the 21st century," 2021, CSO Online, https://www.csoonline.com/article/2130877/the-biggest-data-breaches-of-the-21st-century.html

International Monetary Fund (IMF). 2018a. "Measuring the Digital Economy." IMF Policy Paper, Washington, DC.

International Monetary Fund (IMF). 2018b. "The Bali Fintech Agenda." IMF Policy Paper, Washington, DC.

International Monetary Fund (IMF). 2020. "Digital Money across Borders: Macro-Financial Implications." IMF Policy Paper 2020/50, October. Washington, DC

Jappelli, Tullio, and Marco Pagano. 2002. “Information Sharing, Lending and Defaults: Cross-country Evidence.” *Journal of Banking and Finance* 26 (10): 2017–45.

Jia, Jian, Ginger Zhe Jin, and Liad Wagman. 2021. “The Short-Run Effects of the General Data Protection Regulation on Technology Venture Investment.” *Marketing Science* 40 (4): 661–84.

Johnson, Garrett, Scott Shriver, and Shaoyin Du. 2020. “Consumer Privacy Choice in Online Advertising: Who Opts out and at What Cost to Industry?” *Marketing Science* 39 (1): 33–51.

Johnson, Garrett, Scott Shriver, and Samuel Goldberg. 2020. “Privacy and Market Concentration: Intended and Unintended Consequences of the GDPR.” SSRN Working Paper 3477686.

Jones, Charles I., and Christopher Tonetti. 2020. “Nonrivalry and the Economics of Data.” *American Economic Review* 110 (9): 2819–58.

Kashyap, Anil, and Anne Wetherilt. 2019. “Some Principles for Regulating Cyber Risk.” *AEA Papers and Proceedings* 109:482–87.

Kopp, Emanuel, Lincoln Kaffenberger, and Christopher Wilson. 2017. “Cyber Risk, Market Failures, and Financial Stability.” IMF Working Paper 17/185. International Monetary Fund, Washington, DC.

Lewis, Randall, and Justin Rao. 2015. “The Unfavorable Economics of Measuring the Returns to Advertising.” *Quarterly Journal of Economics* 130 (4): 1941–73.

Marotta, Veronica, Vibhanshu Abhishek, and Alessandro Acquisti. 2019. “Online Tracking and Publishers’ Revenues: An Empirical Analysis.” Unpublished.

Mattoo, Aaditya, and Joshua Meltzer. 2018. “International Data Flows and Privacy: The Conflict and Its Resolution.” *Journal of International Economic Law* 21 (4): 769–89.

Meltzer, Joshua P. 2020. “The Court of Justice of the European Union in Schrems II: The Impact of GDPR on Data Flows and National Security.” Research report, Brookings Institution, Washington, DC.

Menon, Ravi. 2018. Presentation at Singapore FinTech Festival 2018, Singapore, November 12.

Nissenbaum, Helen. 2009. *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Stanford, CA: Stanford University Press.

Peloquin, David, Michael DiMaio, Barbara Bierer, and Mark Barnes. 2020. "Disruptive and Avoidable: GDPR Challenges to Secondary Research Uses of Data." *European Journal of Human Genetics* 28:697–705.

Pierri, Nicola, and Yannick Timmer. 2020. "Tech in Fin before FinTech: Blessing or Curse for Financial Stability?" IMF Working Paper 20/14, International Monetary Fund, Washington, DC.

Sahay, Ratna, Ulric Eriksson von Allmen, Amina Lahreche, Purva Khera, Sumiko Ogawa, Majid Bazarbash, and Kimberly Beaton. 2020. "The Promise of Fintech: Financial Inclusion in the Post COVID-19 Era." Departmental Paper 20/09, International Monetary Fund, Washington, DC.

Stiglitz, Joseph, and Andrew Weiss. 1981. "Credit Rationing in Markets with Imperfect Information." *American Economic Review* 71 (3): 393–410.

Su, Jessiva, Ansh Shukla, Sharad Goel, and Arvind Narayanan. 2017. "De-anonymizing Web Browsing Data with Social Networks." *Proceedings of the 26th International Conference on World Wide Web.* ACM Digital Library.

Sun, Tao, Alan Feng, Yiyao Wang, and Chun Chang. 2021. "Digital Banking Support to Small Businesses amid COVID-19: Evidence from China." IMF Global Financial Stability Notes 2021/02. International Monetary Fund, Washington, DC.

Veldkamp, Laura, and Cindy Chung. 2019. "Data and the Aggregate Economy." Working paper, Columbia University, New York.

Vives, Xavier. 2016. *Competition and Stability in Banking: The Role of Regulation and Competition Policy*. Princeton, NJ: Princeton University Press.

Winegar, A. G., and C. R. Sunstein. 2019. “How Much Is Data Privacy Worth? A Preliminary Investigation,” *Journal of Consumer Policy* 42 (3): 425–40.

World Bank. 2021. *World Development Report 2021: Data for Better Lives*. Washington, DC.

World Economic Forum. 2018. *The Appropriate Use of Customer Data in Financial Services*. White paper, Geneva.

Annex I. What Is Data?

Data on individuals has long been used in commerce, finance, and public policy, but it has proliferated with digitalization.[24] From ancient Mesopotamia to post-enlightenment censuses, states and private entities have sought data on individuals with growing reach and detail. Two recent technological trends have led to an explosion in the economic relevance of data. First, technological progress has drastically reduced the costs of collecting and storing data. Widespread digitalization leads to more data being produced as a by-product of economic and social activities. Second, advances in analytic techniques have allowed for more sophisticated processing to extract greater value from available data. General-purpose technologies, including artificial intelligence and machine learning, have pushed the use of massive databases across sectors, with prediction algorithms widely deployed to identify promising new drugs, deliver financial services to previously excluded households and small and medium enterprises, deliver targeted advertising, and improve the efficiency of operations.

This Staff Discussion Note focuses on individual data, which can be mapped to people's attributes and behavior. The salient feature is that the information is always at the level of an individual, unlike other types of data used in macroeconomics and finance that can be aggregated at levels of entities (sectors, companies) or geography (regions or states). Such data can reflect physical attributes and behavior of human beings (such as gender and age), economic characteristics (including income, property, and transactions), social connections (friends, professional networks, and so on), tastes (as reflected for instance in web browsing habits and purchase history), and sensitive personal data (such as health characteristics and security information). The individual whose data is being recorded can in many cases be identified or geolocated. The incredible breadth and detail of direct and incidental information gathering now feasible, coupled with unclear agency by individuals over control of their data, is at the heart of privacy challenges but also opportunities in commerce and finance that we explain.

Individual data provide new opportunities and challenges for the measurement of the economy. As discussed in Hammer, Kostroch, and Quirós-Romero (2017), the proliferation of digital capture of individual data can notably improve economic measurement through three main channels. Measurement

[24] Previous IMF work has studied the implications of big data and digitalization for the compilation of economic statistics (IMF 2018a) and for real-time policymaking (Hammer, Kostroch, and Quirós-Romero 2017).

improves first by answering new questions and producing new indicators; second by bridging time lags in the availability of official statistics and supporting the timelier forecasting of existing indicators; and third, as an innovative data source in the production of official statistics. However, the incorporation of individual data into reliable economic indicators fit for policy use presents some challenges, including data quality, difficulties with access, and new skills and technologies accessible also to less developed countries. International statistical cooperation is key to overcoming big data challenges, including in the context of the important ongoing FSB-IMF G20 Data Gaps Initiative (DGI).[25]

[25] The initiative has recommended "data sharing" both within and across countries. Progress has so far been limited, reflecting the sensitivity of individual and firm-level data.

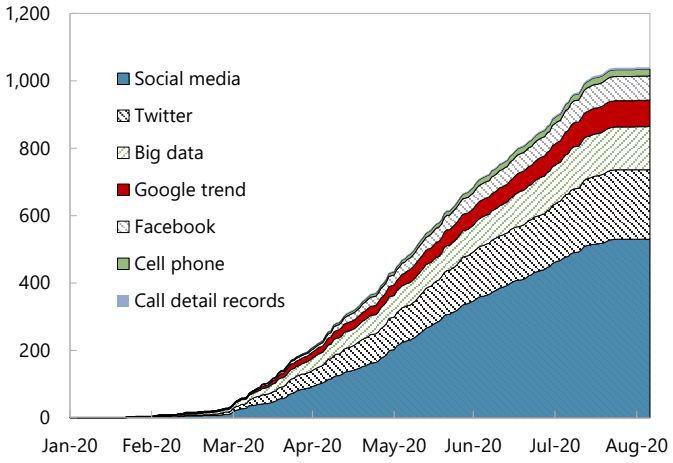
Appendix I. Background Boxes

Box A1. Data and the Pandemic

During the pandemic there has been a surge in the use of big individual data sets to analyze the spread of the virus and the effect of policies. Real-time big data from various platforms has been used for contact tracing and mobility tracking. This is captured in the jump in policy analysis using cell phone data and information from social media and other platforms (World Bank 2021).

Published Articles by Type of Private Intent Data Used

(Number of articles)



Source: COVID-19 Open Research Dataset; and World Bank (2021).

Privacy protections have at times come into conflict with necessary cross-border medical data sharing. Differing privacy protection standards have made it harder to collaborate on crucial medical research across borders because of the difficulty of sharing individual results of biomedical trials (Peloquin and others 2020). Solutions to such tensions are even more pressing in the context of the pandemic.

Box A2. Open Banking

Open banking initiatives promote competition and innovation through data sharing in the financial sector. These policies have been implemented in many countries, including Australia, Brazil, the European Union, India, Mexico, Singapore, and the United Kingdom. Open banking can be thought of as a data access policy for the financial sector that allows for data sharing subject to consumer consent (Carrière-Swallow and Haksar 2021a). These policies aim to change how data—and thus information—flows in the financial system: who has it, who doesn't, and who decides. They recognize that sharing of data across incumbent and potential entrants in the financial services sector can facilitate entry, competition, and innovation through new and better products and services.

Open banking frameworks vary across jurisdictions. They range from public mandates for reciprocal data sharing among all regulated entities at the initiation of the consumer, to public encouragement by regulators, to private-sector-led initiatives with public neutrality. Data sharing is often facilitated by the development of open application programming interfaces (APIs) that allow data to be transferred securely in standardized formats. Open banking involves several innovations with respect to information sharing that takes place through traditional credit bureaus. First, financial institutions exchange data about customers with each other directly, rather than doing so through an intermediary. This allows them not only to obtain a processed credit score, but also to use the granular data to do proprietary analysis and offer more customized products. Second, users are given more control over their data in the open banking model. Although credit bureaus tend to authorize data transfer when the customer engages in certain predetermined tasks—for instance, submitting a rental application—open banking envisions that the user can initiate a data transfer at will and determine what is shared with whom.

The success of open banking may depend on integration with other public infrastructures and policy supports. Provision of digital ID and setting standards for the interoperability of payments are two key aspects, which feature prominently in India's open banking infrastructure (Carrière-Swallow, Haksar, and Patnam 2021). A national system of digital identity can lower the cost of identity verification and facilitate compliance with money laundering and terrorism financing requirements by strengthening the capacity to implement know-your-customer standards. Around the world there has been concern with the scale of digital ID provision under the auspices of central governments and the potential for its use in applications that infringe on individual rights for privacy. This suggests that to successfully implement such a stack-based approach, there is a need for a modernized privacy framework.

Box A3. National Approaches to Data Privacy Frameworks

Privacy: Some frameworks define privacy in constitutional terms as a fundamental right and freedom. Others frame the right as a common law concept that protects data in commercial and financial transactions. One important difference is with respect to the role of the state. Many have exceptions to data privacy when a clear public interest can be determined—for example, to support regulatory requirements or law enforcement actions.

Spillovers: Many frameworks have extraterritorial reach, offering data protection outside their home jurisdiction. The General Data Protection Regulation (GDPR) covers not only European entities but also data processors outside the European Union that process personal data for goods and services offered in the EU or for monitoring the behavior of individuals in the EU.

Consent: The establishment of lawful consent to process individual data varies across countries. In most frameworks, consent to data processing for defined legal purposes must be freely given and explicit. There are exceptions, and most frameworks also allow for exemptions if a legitimate interest can be demonstrated.

Portability: This is a right that allows data subjects to obtain their personal data and to reuse it for their own purposes. To encourage competition, many frameworks require that personal data be provided to the data subject in a structured, commonly used, and machine-readable format and that it can be transferred to a different data controller, data fiduciary, or service provider. There are some differences as to whether the data subject can mandate data transfer to a third party and over the perimeter of information that must be made portable.

Localization: Data localization laws require that data be stored or processed within the country. These policies seek to assert countries' data sovereignty and protect sensitive data from misuse and cybersecurity threats. Some governments have introduced data localization laws that either require or encourage companies to store individual data on their citizens within national borders or—in some instances—restrict the transfer of individual data across borders. Others establish extraterritorial jurisdiction over foreign businesses whose activities have the potential to harm national security and explicitly allow for the use of localization policies to retaliate against discriminatory technology trade and investment measures imposed abroad.

Security requirements mandate that data be processed in a secure way. This includes protection against unauthorized or unlawful processing and against accidental loss or damage. Some frameworks do not impose data security requirements but include a right of action for certain data breaches that result from violations of a business's duty to adhere to reasonable security practices and procedures in a risk-based manner.

Penalties: These vary across jurisdictions in scope and size. The GDPR imposes monetary penalties for noncompliance. Penalties range from 2 to 4 percent of global turnover, depending on the severity of the violation, and are relatively high compared with those in other jurisdictions.

Box A4. Data as Property

Heavy debate: Treating data as property, whether as real property or something more akin to intellectual property, is a recurring idea that has been heavily debated. The basic premise is that data has value, which firms are willing to pay for, and that often this data is about, generated by, or related to individuals and, as such, they should have property rights over it (for example, the right to exclude others or to charge accordingly). The overall concept has not been adopted widely, though some regulatory regimes have property-like aspects.

Complications: Constitutional questions arise in some jurisdictions about whether individuals can assign away their right to data privacy. Data as property could lead to inequities. For instance, people with less wealth (and more willing to sell their data) will have less privacy. On the other hand, not considering the property aspect can have unexpected consequences, such as failing to address property rights over the data of the deceased.

Fragmentation: Countries have different views of property rights for data, particularly when these are weighed against other objectives, such as privacy or national security. Diverging views could lead to data processors shifting activities to jurisdictions with fewer protections (for example, raising equity concerns such as “privacy for sale”).

Options: There continues to be valuable discussion on whether certain aspects of property law—intellectual property law in particular—may be references for setting up data regulation (for example, who has the rights to revenue streams arising from the use of individual data).