Data Financing for Global Good

A Feasibility Study

| Vili Lehdonvirta | Brent Daniel Mittelstadt | Greg Taylor |
| Yin Yin Lu | Artem Kadikov | Helen Margetts |

oiioiioi
oiioiioi
oiioiioi
UNIVERSITY OF OXFORD

The ROCKEFELLER FOUNDATION

This publication was based on research funded by the Rockefeller Foundation.
Grant reference (2015 INF 313).

Suggested citation: Lehdonvirta, V., Mittelstadt, B. D., Taylor, G., Lu, Y. Y., Kadikov, A., & Margetts, H. (2016) *Data Financing for Global Good: A Feasibility Study.* Oxford: Oxford Internet Institute.

Pēteris Zilgalvis, J.D.
Head of Unit, Start-ups and Innovation
Digital Single Market Directorate
Directorate General Communications Networks,
Content and Technology European Commission

FOREWORD

Data are being compared to many things: oil, commons, infrastructure, toxic waste, as well as other beasts, liquids, structures, vegetables and combustibles. What we can be sure of is that all this attention is a sign that the role of data in our society, economy and Digital Single Market is a central one and will only grow in importance.

Even now, data disrupts our social models, civil society, regulatory frameworks and business models. How data are collected, processed and stored, by whom, who has access to it or controls or 'owns' it, and who is liable if it is corrupted are key determinants of whether models for business and social innovation succeed or not. Changes to the tax base in many countries due to the transformation brought about by globalisation, labour saving new technologies and the 'sharing economy' are similarly disruptive.

Alongside these challenges of the 'data economy', new opportunities are created to involve data intensive industries in financing socially beneficial projects, such as a 'data extraction tax' to fight extreme poverty. The Rockefeller Foundation has taken these important opportunities under serious consideration, culminating with a study on the feasibility of innovative financing for social good in the data economy. This report addresses the findings of that study.

We often hear the aside about policy-based evidence making: the purported case where a decision has already been made by a politician to undertake a given initiative, with evidence cobbled together after the fact to justify the choice and disqualify other courses of action. This study aims for precisely the opposite, by providing objective analysis and an invaluable evidence base for policy makers and citizens considering whether and how to start developing innovative financing mechanisms in the data economy. I trust you will find it achieves this aim.

TABLE OF CONTENTS

EXECUTIVE SUMMARY

The commercial world is undergoing a broad transformation associated with recent technological developments that generate unprecedented volumes and types of data. This so-called 'data economy' thrives on the generation of value from analysis of data streams describing social and individual phenomena. However, the explosion of value that has been set off comes with its own set of social challenges, such as information security and personal privacy. Moreover, large sections of global society have yet to experience the benefits brought by these transformations.

The industries that have capitalised on the data economy are in a position to contribute to global good, by investing in infrastructure, curbing negative externalities and sharing benefits with wider stakeholders, while also helping to protect stakeholders' rights and interests. Estimates place the value of the data economy in rich industrialised countries at around 7 percent of gross domestic product, or approximately ten times the value of the United Nations development aid spending goal, suggesting that the impact of such contributions could be significant. A data-linked levy or other innovative financing mechanism applied to the data economy could be used to realise such contributions in practice. The range of potential data financing mechanisms can be illustrated with four models:

1. A global internet subsidy

One broad class of justifications for data financing concern the need to provide public goods and correct for positive externalities such as network effects in technology adoption through subsidization schemes. For example, subsidization of Internet access for the poorest members of global society would not only provide access to them, but also benefit other users and online service providers by growing the population of Internet users. More broadly, there are many prerequisites for technology adoption (such as good education) that could be funded under such a scheme. This suggests that a global fund to make strategic investments in development, with a focus on technology and connectivity would be of significant value. Such a fund could be used in several ways to encourage Internet uptake, from funding local start-ups or ISPs to building infrastructure on either proprietary or open technological protocols. A subsidy for global Internet access could be funded in several ways. Online companies could voluntarily contribute to the fund—perhaps according to an agreed metric, e.g. number of users or revenue generated from secondary uses or sales of user data. Alternatively, a data financing mechanism on Internet users could be justified by the subsidy's reduction of 'digital divides' between online and offline individuals, groups and regions.

2. A 'privacy insurance' for personal data processing

Processing vast amounts of personal data can be very valuable to business, but it also involves an inherent risk of data leaks. Data leaks can cause harm to data subjects through privacy loss and identity theft as well as to wider society and industry by enabling unauthorised access and ultimately reducing public trust. It could be proposed that responsible data processors should address this possibility pre-emptively, by contributing to a common fund. The fund could be spent on the development of privacy-enhancing technologies, or compensation for victims, thereby providing contributors with a kind of insurance against the costs associated with data loss. Precedent exists for mandating the uptake of insurance where actors take on risks that may affect others—for example, motorists are required to have insurance for third-party damaged and many businesses in some jurisdictions are required to have employer's liability insurance. Alternatively, in recognition of the harm caused to wider society, some fraction of the revenue from this scheme could be diverted for other social responsibility projects that benefit society, such as the development and public good projects identified above. A suitably judged contribution rate could also serve to address the negative externality inherent in data retention by inducing data controllers to be more circumspect about the data that they

hold. The responsible use of such a public insurance scheme might also help data processors to 'insure' themselves against the reputational damaged of a data breach.

3. A 'shared knowledge duty' for open and public data

Wealth is not created in isolation and is dependent on economic and political communities. As part of a social contract with data subjects, data processors generating revenue from shared resources may have a duty to share the benefits of processing fairly. There is a case to be made that contributors to a dataset should be entitled to usage of proprietary information, products and services created from processing of public data, without having to share additional personal data or pay the market rate for access, because the creation of the public dataset already constitutes a type of payment. In other words, exploitation of shared resources would need to benefit the public or groups that contribute to the dataset, which can include society in general or specific groups of users conceived as 'digital labourers'. On this basis, a duty could be levied for revenue-generating processing of open and public data. Such a duty could be applicable to both public datasets that have been funded by public resources or taxation and the development of products or services from the 'free labour' of Internet users (e.g. scraping user content to improve a personalisation algorithm).

4. An 'attention levy' for digital marketing

A significant application for the processing of personal data has been the online advertising industry. Users' attention is often exploited and that this may lead both to economic inefficiency and also to violation of individuals' sovereignty over their personal or private sphere. Both issues could be addressed with a corrective levy on intrusive advertising that compels organisations to be more circumspect about invading a user's attention. A levy on abuse of users' attention could simultaneously protect users from annoying or distracting material while also cultivating an environment in which users are more receptive to non-abusive messages.

Sound economic, social and ethical arguments can be found to justify a variety of data financing models, including the four already mentioned. Investment in (and adoption of) technological infrastructure has positive spillover benefits for stakeholders, while there are negative spillover harms when data processors expose users to privacy breaches or when network users cause congestion. These spillover benefits and harms undermine the market's efficiency, but could be corrected with a judiciously constructed data financing mechanism. Elsewhere, there is a need to protect the autonomy of individuals. However, a substantial body of evidence suggests that people make uninformed and systematically damaging decisions when it comes to divulging private information. With this in mind, a carefully judged data financing scheme could also serve a paternalistic role in protecting users from the harm that such poor choices entail. Thirdly, one can argue that digital businesses occupy a special role in society by virtue of the particular economics of the markets they inhabit, and that they have benefited immensely from exploiting fundamentally public resources. There is, therefore, an ethical argument that financial contributions from digital businesses would ensure they honour their social contract by sharing benefits with society at large. In sum, data financing would not—and should not—simply be about redistribution for the sake of redistribution, but about correcting economic and ethical problems arising in the data economy.

That said, there are also some important pitfalls involved in data financing. If firms are asked to pay a levy and respond by increasing their prices then the ultimate cost may fall upon consumers, whom the levy might be intended to protect. Thus, careful consideration of who will ultimately bear the cost of the financing mechanism is required. Moreover, many people use and benefit from the services offered by digital companies. But taxes distort market allocations in ways that often reduce the opportunities for such beneficial trade and that are detrimental for overall welfare. A third important point is that a levy may reduce innovation by decreasing the rewards it carries.

Depending on what justification is used to create a data financing scheme, a range of different stakeholders may be asked to contribute. For instance, contributors could be some subset of Internet companies, other data-intensive industries, digital advertisers, data brokers, networking hardware vendors, cloud service providers, intermediaries including app stores, the Internet Corporation for Assigned Names and Numbers (ICANN), or even Internet end-users, corporate or individual.

Any innovative financing mechanism also requires a basis for calculating the contributions due from each stakeholder. This is where many of the practical challenges of data financing lie. The most obvious approach is to use the volume or quantity of data processed or stored as a measure on which dues are based. However, neither measure is directly and consistently associated with the value or consequences of data processing. They are also extremely context-dependent measures, with the consequence that the results could be highly arbitrary if applied across a variety of firms and industries. For data financing models that target very specific industries, volume or quantity might offer a feasible basis for calculating dues.

Another approach to calculating dues in a data financing mechanism is to measure the value of data. For certain limited types of data that are bought and sold, it is possible to observe a market value. But much of the data trade happens 'over the counter', with the consequence that transactions are not observable; this is the case with secondary uses of personal data collected from users, for instance. It may also be possible to use information disclosed in company financial reports to produce rough proxies of the value of firms' data-intensive activities. The value of personal data has also been estimated with surveys and economic experiments. Measurement of data and value are not merely line-drawing exercises: consideration must be given to the technical feasibility and operating cost of any measurement system.

Assuming appropriate stakeholders can be identified and value assigned to data or a proxy thereof, political and legal challenges still remain. Two key dimensions for a data financing scheme are its degree of compulsion (e.g. is it a voluntary scheme, or is contribution mandated through the apparatus of taxation) and its degree of multilateralism (e.g. does a single country or a single organization act alone, or in concert with partners internationally). The positioning of a scheme in these dimensions determines the legal and political barriers that must be overcome, and is closely related to the practical and technical challenges of implementation. For instance, if a data financing mechanism is voluntary rather than mandatory, it may be possible to avoid many of the practical and technical challenges, as long as contribution levels can be set in agreement between members. A unilateral or multilateral CSR scheme may be the most politically and legally feasible way of structuring a data financing mechanism. Indeed, many leading information technology companies already contribute to social good through CSR initiatives, though some of these initiatives are also aimed at further entrenching the company's own market position in the global data economy. There might be opportunities to build on these individual initiatives to create multilateral industry-spanning efforts that would deliver social good whilst moderating individual companies' attempts at market dominance.

International political momentum also exists that can be used towards establishing a compulsory data financing model, focused on issues of Internet development and data protection. There are, however, a number of significant political and legal barriers that would have to be overcome before any compulsory data financing model could be introduced. The key obstacles are the lack of a legitimate governance mechanism, differing political and legal agendas across key players in the international arena, as well as the availability of other, possibly more appropriate or efficient legal mechanisms for regulating the data industry. In particular, the importance of a trans-national governance regime—which would administer calculation of tax liabilities, and collection and distribution of funds—should not be underestimated.

Industry attitudes are also an important

contributor to the feasibility of any data financing proposal. Six types of concerns are likely to be prominent in industry responses to any data financing proposals: concerns related to effects on innovation, value for users, ownership of personal data, effects on firms of different sizes, potential for bureaucracy and waste, and principled opposition to any public sector led initiatives. Data financing might be particularly problematic for start-up companies experimenting with technologies and business models. From the industry perspective, it is important to consider how data financing models could contribute value rather than merely extracting it. For instance, duties collected from revenue generated from open datasets could perhaps support the increased availability of such datasets, thus increasing revenues in the long run.

A suggested next step to move data financing from an idea towards implementation is conducting further studies that narrow down the scope by taking a sectoral approach, focusing on a specific industry, vertical, or form of data or data processing. Technical solutions or other practical means (e.g. financial reports) to establish a data financing mechanism in particular sectors can also be examined. Alternatively, further studies could focus on particular legal structures, such as a multilateral consortium of Internet companies, or an international legal instrument building on wider political momentum. Consideration can be given to the activities that can be brought under the umbrella of particular legal structures. Narrowing down the scope would allow studies to examine in detail the applicable subset of the many issues outlined in this report, perhaps developing concrete solutions. Studies could perhaps also approach the topic from an entirely different direction, asking what global problems there are that could best be addressed by data financing, considering the type and scale of problem, and any linkages to the global digital economy.

This report provides a first step in the assessment of data financing as a mechanism for social good in the data economy. To use the language of the technology industry, data financing is a 'moonshot'—a radical idea whose success is far from certain, yet addressing such a huge problem that it is tempting to at least give it a try.

1 INTRODUCTION

The commercial world is undergoing a broad transformation associated with recent technological developments that generate unprecedented volumes and types of data (see: Figure 1). The so-called 'data economy' thrives by unlocking value from the analysis of data streams that describe social and individual phenomena. However, the explosion of value that has been set off comes with its own set of social challenges such as information security and personal privacy. Moreover, large sections of global society have yet to experience the benefits brought by these transformations.

This report assesses the feasibility of data financing: a type of innovative financing intended to capture a small slice of the value created in the commercial exploitation of data and redirect that value towards achievement of broader social objectives.

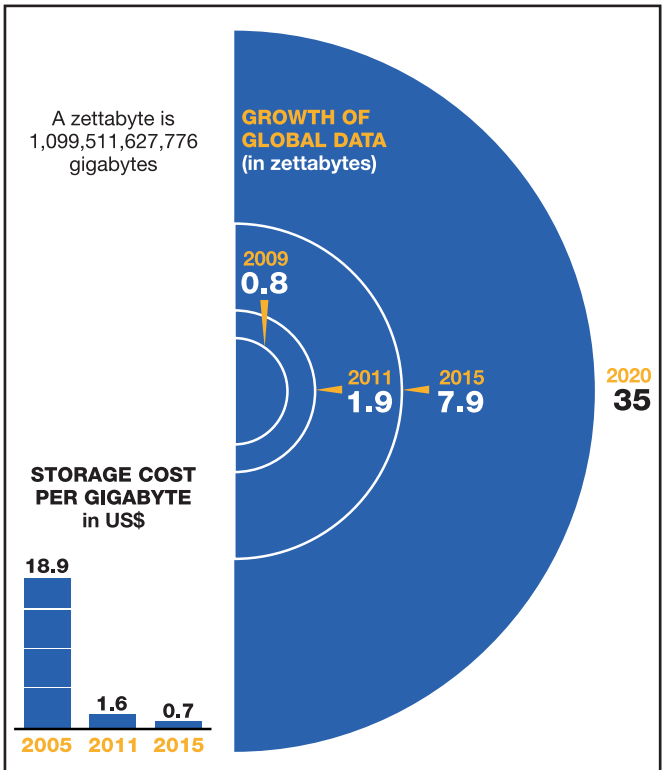


Figure 1: Growth of Global Data

Figure based on estimates from Thomson Reuters, available at: http://blog.thomsonreuters.com/index.php/big-data-graphic-of-the-day/.

Innovative financing is a term that is used to refer to initiatives that seek to unlock private capital for the sake of global development and socially beneficial projects, which face substantial funding gaps globally.[1] This report provides a map of issues, contingencies and opportunities faced when planning innovative financing mechanisms in the data economy, or 'data financing'.

Data financing, like data, is an inescapably transnational concept. As an intangible asset, data can be transferred seamlessly across national borders and jurisdictions. This fluidity provides unprecedented opportunities for companies to create value around the world, but also difficulties for any attempts to 'pin down' data for regulation or taxation. It raises many difficult questions: for instance, who should contribute, based upon which measures, for whose benefit?

The idea of data financing is not entirely without precedent. Lessons can be taken from prior proposals and initiatives such as the UNITAID[2] global health initiative or the OECD's proposed 'bit tax'.[3] Prior attempts to tax e-mail and Internet bandwidth are particularly informative when it comes to potential difficulties: both faced stern opposition in the 1990s and even elicited an explicit ban in the United States in the form of the Internet Tax Freedom Act of 1998. A significant weakness of both proposals was the perceived arbitrariness of the quantity of e-mails and bandwidth as a tax basis. The implementation of any data financing mechanism likewise faces many conceptual and practical difficulties. But if realised responsibly and at scale, the impact on global good could be significant.

This report summarises the results of a study commissioned by the Rockefeller Foundation and conducted by a multidisciplinary team based at the Oxford Internet Institute to investigate the feasibility of data financing. The purpose of the study was to take the

[1] Based on estimates by the Rockefeller Foundation, global philanthropic funding of sustainable development faces a reported $2.5 trillion annual funding gap. Available at: https://www.rockefellerfoundation.org/blog/development-goals-without-money-are-just-a-dream/

[2] Further information on the scope of UNITAID is available at: http://www.unitaid.org/en/

[3] OECD. (2014). *Addressing the Tax Challenges of the Digital Economy*. Paris: Organisation for Economic Co-operation and Development. Retrieved from http://www.oecd-ilibrary.org/content/book/9789264218789-en

first step in assessing whether and how data financing mechanisms might work to address global funding gaps, and what forms they could and should take. The study centred on a series of four expert workshops, bringing together stakeholders from around the world across academia, the public sector, and data-intensive industries. Each workshop examined the feasibility of data financing from a particular angle in order to bring in a wide range of disciplinary and sectoral perspectives. Insights from the workshops were supplemented with desktop research and fact-finding missions focusing on previous proposals and initiatives, legal and conceptual analysis, simple economic estimates, and feedback from stakeholders in data-intensive industries. The result is a wide-ranging map of the issues and challenges involved, by no means exploring each avenue anywhere near to its end, but providing a chart of the general lay of the land.

First, the justifiability of data financing is considered. Any value-generating activity can be taxed to fund valuable social goals, but is there any reason to target data specifically? Could data financing mechanisms not only redistribute value but generate it, or at least address an ethical responsibility specific to the data economy? The strength of the economic, social and ethical foundation upon which a proposed mechanism might be built must be assessed for two reasons. First, all parties concerned will naturally wish to know that such a mechanism does not imply unforeseen social or ethical consequences that might undermine its otherwise well-intentioned objectives. Second, when decision-makers and industry stakeholders are asked to endorse or accept the introduction of such a mechanism, they might be expected to consent more readily if they can be presented with a clear and cogent argument justifying the existence of the mechanism.

Second, the practical and technical implementability of data financing is considered. Much of the difficulty with data financing can be traced to the malleability of 'data' as a concept. Many technical and practical barriers are faced by any mechanism that proposes to use data or data processing as a tax base or measure. An important challenge that must be addressed is the fairly pragmatic issue of determining what should be subject to a data financing mechanism, and when. The ways in which data are defined, held, and used are so diverse that it is essential to define the scope of a scheme clearly. This is not merely a line-drawing exercise: due consideration must also be given to the cost and feasibility of the proposed scope—which, in turn, depend upon factors such as the observability and auditability of transactions, and the technical requirements for monitoring and administering the tax.

Third, the international legal and political dimensions are also essential for determining the feasibility of data financing. Data, like the technology sector that has been so instrumental in its growth, is inherently trans-national. Policymakers already wrestle with jurisdictional issues created by flows of data across legal boundaries. Areas of law such as data protection, privacy, and net neutrality find a new relevancy in the data economy. When designing a data financing mechanism, it is important to be mindful of both the political and the legal framework in which it is to be implemented. Two critical dimensions of any proposed mechanism are the degree of compulsion and centralisation. Mechanisms can range from entirely voluntary charitable contributions, comparable to existing corporate social responsibility (CSR) initiatives, to compulsory taxes. Similarly, mechanisms can range from decentralised schemes administered by individual organisations to multilateral cooperative efforts.

Finally, industry attitudes towards data financing are essential to its success. A data financing mechanism would mean additional costs to firms in some or all data-intensive industries, even if, on some time scale and in some form, they could expect a return from it. Even if as a share of industry revenues the costs would be negligible, it is possible that firms might take a negative stance towards such an initiative. An effective industry association taking an actively hostile stance towards a data financing mechanism could be fatal to an otherwise valuable

project. Conversely, private sector partners and associations who see the value of a mechanism and embrace it could be essential in realising such a mechanism in practice.

The remainder of the report is organized around these four themes. We begin, in Chapter 2, by assessing the economic, social, and ethical justifiability of data financing. We introduce conceptual arguments in favour of and against a data financing mechanism. From this foundation, we identify four broad conceptualizations of a data financing mechanism that, rather than amounting to the arbitrary creation of a tax, build upon sound economic ethical principles. These four models serve as focal point for the discussion that follows. In Chapter 3 we discuss the technical and practical feasibility of implementing a data financing mechanism, while Chapter 4 examines the legal and political challenges inherent in introducing such a scheme. Chapter 5 describes the response of industry stakeholders and outlines some of their key concerns. Chapter 6 concludes and provides rough estimates of the potential scale of the four proposed data financing models.

2 JUSTIFIABILITY OF DATA FINANCING

Key questions

- What justifications can be offered for data financing?
- What are some of the potential pitfalls from such a scheme?
- What are the possible data financing models that could be further developed in practice?

Summary

- A number of social, economic, and ethical justifications for data financing are possible.
- Investment in (and adoption of) technological infrastructure and supporting factors has positive spillover benefits for stakeholders—including digital businesses. A public scheme of taxation and subsidization can help to ensure that public goods are efficiently supplied and positive externalities are realized.
- Negative side effects arise from business activities that arise as a result of exploitative collection, processing or storage of data; of aggressive use of people's attention; or of network congestion. These side effects could be mitigated through a corrective data financing mechanism that deters behaviour with spillover costs.
- While there is a need to protect the autonomy of individuals, there is evidence that people make uninformed and systematically damaging decisions when it comes to divulging private information. There is, therefore, potentially scope for a paternalistic data financing mechanism that would both deter data processors from exploitation and provide an indirect form of compensation for the associated harms.
- Financial contributions from digital businesses could also ensure that they honour their social contract by sharing more of the benefits they earn through exploitation of public resources (such as publicly-funded open data) or through appropriation of free labour.
- Society might legitimately expect compensation when firms achieve (and are permitted to exploit) a position of dominance thanks not to their own merits, but rather because of the intrinsic characteristics of the market in which they operate.
- A number of pitfalls associated with data financing are identified.
- If firms are subject to taxation, then they are likely to pass at least some of the cost onto consumers. It is important to be mindful of who will ultimately bear the burden of any data financing mechanism.
- Many people use and benefit from the services offered by digital companies. But taxes distort market allocations in ways that often reduce the opportunities for such beneficial trade and that are detrimental for overall welfare.
- A data financing mechanism that reduces the profitability of investments in new infrastructure or in innovation is likely to reduce the incentive to engage in such investments, with potentially significant long-run costs to society.
- We propose four possible models of data financing for further consideration. These models comprise a global Internet subsidy; a privacy insurance; a shared-knowledge duty; and an attention levy.

Analysis

This chapter proceeds as follows: a range of conceptual justifications are first identified that might be used to argue in favour of data financing. The purpose here is to examine whether a data financing scheme can go beyond providing resources for philanthropic projects, and also serve some broader social goal—such as correcting for pre-existing distortions or inequities. These justifications are then contrasted with important objections to data financing: it is likely to distort decisions it a manner harmful for the efficient functioning of markets, and may

well discourage investment and innovation. Four potential models for a data financing scheme are then introduced. These provide a focal point for the remainder of the report, which explores concepts and challenges demonstrated by each model.

2.1 Economic, social, and ethical justifications for data financing

2.1.1 Contribution to public goods

A historically important motivation for taxes and levies has been the need to finance public goods. A good is said to be public if it can be enjoyed by many consumers simultaneously (non-rivalry) and if there is no way to prevent any one individual from consuming it (non-excludability). It is well-known that markets typically fail to provide such goods efficiently because everyone has an incentive to free-ride: instead of contributing to the cost of the public good themselves, they wait for others to do so in the hope of consuming it for free. The truly perverse thing about this kind of market failure is that everyone can agree that the good is being under-provided and yet still be unwilling to increase their own contribution. Authorities such as governments have taken a leading role in resolving this collective action problem. Instead of counting on individual actors to unilaterally provide the good (inevitably leading to free-riding), the goods are provided through a central mechanism of taxation and fiscal expenditure that demands that all contribute and allows all to benefit.

One feature of digital services is that it is difficult to unambiguously pin-down the location at which a transaction takes place. This can prove problematic for the enforcement of traditional taxation regimes, which are tied to geographic boundaries and play a key role in financing public goods. Indeed, a number of technology companies have been subjected to intense criticism for shifting their profits across jurisdictions in a way that results in minimal tax obligations.[4,5] For example, Facebook UK's accounts record 2014 revenues of £105 million ($140 million), yet the relocation of profits overseas meant that the company paid only £4,327 ($5,644) in UK corporation tax.[6] Google, Amazon, and others have been criticized for similar arrangements. The objection is not a legal one: such accounting schemes are carefully crafted so as to be in full compliance with the relevant laws. Rather, there is a robust argument that organizations conducting a large volume of business within a jurisdiction (and thereby consuming the benefits of public goods such as protection of their premises by police or the enforcement of contracts by public institutions) should be required to make a financial contribution commensurate with the collective resolution of the public goods problem. This issue was officially recognized in the UK, which introduced a Diverted Profits Tax into legislation in the Finance Act 2015. The OECD has also undertaken work to eliminate tax loopholes.[7] The further propagation of such schemes has the potential to raise significant revenues for publicly beneficial projects, and can be justified as fair compensation for public good consumption.

One feature of digital services is that it is difficult to unambiguously pin-down the location at which a transaction takes place.

Some caution is required in applying the public goods argument. In particular, not all publicly provided goods and services are public goods. Many types of telecommunications infrastructure, for example, are excludable (it is possible to prevent consumption by those who have

[4] Stewart, H. (2015). Facebook paid £4,327 corporation tax despite £35m staff bonuses. *The Guardian*, 11 October. https://www.theguardian.com/global/2015/oct/11/facebook-paid-4327-corporation-tax-despite-35-million-staff-bonuses.

[5] Frankfurter Allgemeine Zeitung. (2016). Google schleuste Milliarden durch Steuerschlupflöcher, 19 February. http://www.faz.net/aktuell/wirtschaft/recht-steuern/google-schleuste-milliarden-durch-steuerschlupfloecher-14079856.html.

[6] Facebook UK Limited (2014) Annual Report and Financial Statements for the Year Ended 21 December 2014. London, UK: Companies House.

[7] OECD (2015) G20 leaders endorse OECD measures to crack down on tax loopholes, reaffirm its role in ensuring strong, sustainable and inclusive growth. https://www.oecd.org/g20/summits/antalya/g20-leaders-endorse-oecd-measures-to-crackdown-on-tax-loopholes-reaffirm-its-role-in-ensuring-strong-sustainable-and-inclusive-growth.htm

not paid). Once consumption can be made contingent on contribution, the possibility of free-riding vanishes and the above logic of under-provision no longer applies. Though consumption of these goods does not intrinsically exhibit public good properties, a free-riding problem may nevertheless exist on the production side. Indeed, online service providers such as search engines, content platforms, or social networking sites "consume" infrastructure every time they transmit content to a user over the network. The principle of net neutrality ensures that, no matter who funded development of the network, anybody can transmit content over segments of the network built by others, thereby free-riding. This is particularly pertinent in developing countries where large, as yet unconnected populations could be reached if only the necessary infrastructure and other enabling factors were in place. Recognizing this opportunity, large online service companies such as Google and Facebook have invested heavily in infrastructure projects. But there is an obvious free-rider problem to the extent that other firms can benefit from use of the infrastructure provided under these schemes.

A centralized scheme that collects contributions more equitably may help to resolve this collective action problem and increase overall levels of investment.

This problem is made more acute by the fact that free-riders may be in direct competition with those who are footing the bill. Attempts to exclude such competitors from the network have met with stiff resistance.[8] A centralized scheme that collects contributions more equitably may help to resolve this collective action problem and increase overall levels of investment.

2.1.2 Creating positive externalities

Closely related to public goods is the concept of a positive externality—a term for spillover benefits from consumption that arise indirectly. For example, education is excludable (and, arguably, rivalrous) and therefore does not fit the definition of a public good. Nevertheless, when an individual becomes educated they create wider benefits in society. Indeed, having a well-educated and technically literate population is recognised as a key precursor to the broad adoption of technology and the widespread benefits that flow from it.[9] Similarly, infrastructure is subject to so-called network externalities: users directly benefit from connecting to the network, but there are also indirect benefits accruing to others because communications technologies become more valuable the more people they can be used to reach. Moreover, other positive externalities flow from technology adoption, such as increased efficiency of public administration.

The interesting consequence of a positive externality is that individuals may conclude that the benefits of consumption are not sufficient to justify incurring the cost, even though consumption *is* worthwhile once the broader social spillover benefits are taken into account. There is thus a tendency for under-consumption, which can be corrected by subsidizing activities that have positive spillover benefits. Like public goods, positive externalities therefore give rise to market failure and provide a rationale for intervention. Technology companies, in particular, look well-placed to benefit from wider adoption of technology and from factors that facilitate this and should have an interest in resolving the associated externalities. A data financing mechanism that promotes adoption of data-generating technologies would therefore provide a legitimate incentive for data-intensive firms to contribute.

2.1.3 Correcting negative externalities

Just as some goods create positive spillovers in consumption, others have negative side-effects (or "negative externalities"). For example, driving a car to the supermarket is convenient, but harms others because it

[8] For example, Facebook's attempts to create a walled garden within its Free Basics Internet service in India came unstuck after intervention from the Indian Telecommunications regulator.

[9] Van Dijk, J. A. G. M. (2005) *The deepening divide: Inequality in the information society*. London, UK: Sage.

creates pollution and congestion. This causes problems when an economic actor perceives an overall private benefit to some activity even though the negative spillovers imply a net cost for society at large. As a consequence, actors will tend to engage too much in activities with negative externalities. A common solution to this problem is to tax the activity in question (e.g., imposing a fuel duty): the tax increases the private cost of the activity to be more in line with the true social cost, and therefore induces people to consider these broader costs when making decisions.

The handling and storage of personal data is a likely source of negative externalities. For a data controller, the costs of storing data are relatively low, and the pay-off to having large amounts of data at hand are potentially large. This provides incentives for data maximalism: a tendency to collect and store as much data about subjects as possible. But there are broader social costs to such data maximalism. As more organisations store more information about each subject, the risk of being the victim of a data leak becomes greater, the harm from such a leak becomes more severe, and the individual's sense of privacy is eroded. Moreover, there are externalities in the protection of data against theft since such protection is often the responsibility of a controller who nevertheless bears only part of the costs associated with a breach. Overall, this suggests that controllers will tend to collect and store too much information (just as individual users tend to on services with free storage). For example, there will be no disincentives to duplicate large-scale data sources rather than implement a more complex but parsimonious or 'data efficient' system. A good example of the latter is the Estonian government's 'X-Road' system (which is used by government, banks and many private organisations), which matches an individual citizen's personal identification number with data registries, meaning that any piece of personal information is stored only once, thereby minimising the quantity of data that is stored and maximising security. Such tendencies towards data maximalism might be corrected with a data financing mechanism on collection and storage of personal data. Presenting such a data financing mechanism as 'insurance' would make a more direct connection between data maximalism and the risk of loss or security breach, and engender more incentives for data efficiency.

Another source of negative externalities is congestion: heavy network usage has a negative effect on the quality of service enjoyed by others. This issue is becoming acute as a relatively small number of major online content providers are accounting for an increasingly large share of the total volume of data transmitted over the network,[10] with a material effect on overall network performance. It was noted above that free-riding in the provision of new infrastructure investment is likely to be an issue in this market. But a congestion externality implies that, even if one were to forget about investment and to take the network infrastructure as fixed, the market would fail to deliver efficient allocation of the available capacity. One means of addressing this issue is, again, to correct the externality with a data financing mechanism (on bandwidth use) similar to the bit tax that was proposed in the 1990s.[11]

For a data controller, the costs of storing data are relatively low, and the pay-off to having large amounts of data at hand are potentially large. This provides incentives for data maximalism: a tendency to collect and store as much data about subjects as possible.

[10] Spangler, T. (2015) Netflix Bandwidth Usage Climbs to Nearly 37% of Internet Traffic at Peak Hours, *Variety* 28 May. http://variety.com/2015/digital/news/netflix-bandwidth-usage-Internet-traffic-1201507187/

[11] It should be noted that such a tax would currently be illegal in the United States under the terms of the Internet Tax Freedom Act of 1998, which prohibits discriminatory Internet-only taxes, including explicitly any type of 'bit' or 'bandwidth' tax. We return to issues of legal feasibility in Chapter 4.

2.1.4 The attention commons and digital trespass

Fish stocks are often subject to over-fishing because each fisherman privately profits from a large catch and has little concern for the sustainability of the stock for others. This is a classic example of 'the tragedy of the commons': when many actors share a common, exhaustible resource, they tend to race to benefit from it while there's still some left—with the upshot that the resource is quickly depleted for everyone. Perversely, it will often be the case that all parties would benefit from agreeing to behave in a more sustainable fashion. But such agreements are difficult to enforce because each party has an incentive to cheat and claim more than his or her fair share.

It has been understood since at least the early 1970s that attention is a scarce resource owing to individuals' limited capacity to process information.[12] Moreover, there is fierce competition for access to this resource among digital advertisers and others with messages they would like to have seen or heard.[13] There is a temptation for unscrupulous organisations to force their way into users' attention, giving rise to phenomena such as excessively intrusive advertisements or spam mail. The resulting information overload not only harms the individual concerned, but also crowds-out more responsible senders who struggle to have their message heard (thanks, for example, to preventative measures such as the growing use of ad blockers). Just as fish stocks are subject to overfishing, the over-exploitation of users' attention has the potential to compromise the sustainability and overall vitality of the online advertising industry. There seems, therefore, scope for both users and responsible businesses to benefit from an initiative that leads to a more sustainable use of the finite pool of attention available. In this spirit, it has been argued that a data financing mechanism on intrusions into users' attention (for example, an online advertising tax, or a very small charge for sending emails) would help to attain such sustainability, and might therefore yield widespread benefits even after accounting for the cost of paying for the mechanism.[14]

Just as fish stocks are subject to overfishing, the over-exploitation of users' attention has the potential to compromise the sustainability and overall vitality of the online advertising industry.

Aside from the tragedy of the commons, there is also an ethical argument to be made here. One could take the view that an individual's attention forms part of their private sphere and that uninvited intrusion on that attention represents a form of trespass, which is inherently illegitimate. More broadly, it could be argued that by using personal data, businesses exploit users' 'digital body'[15] in the same sense that medical research exploits data relating to health. Data-heavy businesses and those that encroach on users' attention could therefore be seen to have an obligation to repay affected individuals or society, which could be achieved through some form of data financing.

2.1.5 Paternalistic protection of Internet users

Users exchange personal data for access to valuable goods and services (such as social networking or search services). A first analysis might suggest that, like a typical economic transaction, such trade is voluntary and therefore must benefit the user providing data. However, it is now well-understood that individuals are poor at rationally judging their interests when it comes to giving

[12] Simon, H. A. (1971) Designing Organisations for an Information-Rich World in *Computers, Communication, and the Public Interest*, ed. by M. Greenberger, pp. 32–72. Baltimore, Maryland: Johns Hopkins Press.

[13] Wu, T. (2016). Attention Brokers. Presented at the Innovation Policy Colloquium, NYU. Retrieved from http://www.law.nyu.edu/sites/default/files/upload_documents/Tim%20Wu%20-%20Attention%20Brokers.pdf

[14] Van Zandt, T. (2004) Information Overload in a Network of Targeted Communication. RAND *Journal of Economics* 35(3): 542–560.

[15] Knight, A., & Saxby, S. (2014). Identity Crisis: Global Challenges of Identity Protection in a Networked World. *Computer Law & Security Review, 30*(6), 617–632.

away personal data. Experimental results show that consumers' willingness to hand over personal data is subject to a variety of behavioural biases that undermine sound decision making.[16] Moreover, consumers are generally not well-informed about their rights in commercial relations and are ignorant of terms of service that they are frequently asked to accept.[17] Therefore, although users provide some form of consent, typically in the form of a service agreement, one should question how meaningful this is in practical terms. A regular user is not likely to be completely aware of what he or she is consenting to, especially when the mere collection of raw data might be followed by complex aggregation and analysis that is beyond users' control, and attributes new and unexpected meaning to the data.[18] Equally important: even without consent, a person's profile can be inferred from various sources of data collected from other users in the same target group.[19] This issue can be particularly pertinent in relation to data about minors, who cannot legally provide consent but whose profiles may nonetheless be inferred from data about others.

Consumers' willingness to hand over personal data is subject to a variety of behavioural biases that undermine sound decision making.

There is therefore a plausible role for paternalism in protecting users against exploitation that takes advantage of their inability to comprehend the nature of the agreement they are forming, or to arrive at a rational judgement about the consequences of that agreement. A data financing mechanism could adjust the imbalance in power, serving as compensation to users for their having provided poorly informed consent to exploitation of their data.

2.1.6 Fulfilment of a social contract

Data-intensive businesses generate value from shared resources and open data, including user-generated content. For example, online search engines can be viewed as leveraging the accumulated knowledge of humankind. Google, for instance, uses officially translated EU documents (that were publicly funded) to teach and refine the Google Translate algorithms using machine learning. These algorithms crowd-out workers and displace translators. We earlier made the economic argument that free-riding on the provision of public goods by others may leave markets unable to function properly. But there is also an ethical argument to be made here based on an implied social contract between data authors and data processors. As data companies have built their business models on public data, there is a case for sharing at least part of the value of that business with the society that has financially supported or helped create that data. A comparison can be drawn with the pharmaceutical industry, which has a history of commercializing and exploiting indigenous medical knowledge, often without feeding any of the benefits back to the societies that initially produced the knowledge.[20] A mechanism based on this justification would need to overcome the objection that services and platforms provided 'freely' by data processors represent sufficient repayment of the social contract.

More generally, a data duty could be used to facilitate distribution of wealth according to norms of social justice.[21] Some have enjoyed immense private gains from technologies

[16] Acquisti, A. and Grossklags, J. (2007) What can Behavioral Economics Teach us About Privacy? *in Digital Privacy Theory, Technologies, and Practices,* ed. by S. di Vimercati, S. Gritzalis , C. Lambrinoudakis and A. Acquisti, pp. 363–377. Boca Raton, Florida: Auerbach Publications.

[17] Rainie, L. (2016, January 20). The State of Privacy in America: What We Learned. Retrieved from http://www.pewresearch.org/fact-tank/2016/01/20/the-state-of-privacy-in-america/

[18] Meaningful Consent in the Digital Economy. (n.d.). Retrieved 4 August 2016, from http://www.meaningfulconsent.org/; Mittelstadt, B. D., & Floridi, L. (2016). The Ethics of Big Data: Current and Foreseeable Issues in Biomedical Contexts. *Science and Engineering Ethics, 22*(2), 303–341. http://doi.org/10.1007/s11948-015-9652-2

[19] Hildebrandt, M. & Gutwirth, S. (2008). *Profiling the European citizen*. Springer.

[20] Ismail, Z., & Fakir, T. (2004). Trademarks or Trade Barriers? Indigenous Knowledge and the Flaws in the Global IPR System. *International Journal of Social Economics*, 31(1/2), 173–194.

that are completely unavailable to the poorest members of global society. It can be argued that these beneficiaries have a duty to contribute in a fashion that sees the benefits of technology distributed more widely. A communitarian justification for a data financing mechanism also fits into this category, as wealth is not created in isolation and is dependent on economic and political communities that require support to be sustained.[22]

2.1.7 Exploitation of free labour

On a closely related note, businesses in the data economy have appropriated not only publicly provided information, but also labour that is essentially provided for free by Internet users. Online product reviews, blog posts or home-made videos, for instance, are often provided by individuals primarily out of an altruistic desire to benefit fellow users or simply to share thoughts, rather than for commercial ends.[23] Nevertheless, these fragments of content have been incorporated as key ingredients for the business models of tech firms and, in aggregate, become extremely valuable.

Often, users may fail to realize that they are providing labour for a firm, or to understand the commercial value of their efforts. For example, many websites determine what content to promote to positions of prominence by monitoring clicking behaviour in a manner transparent to the user. Thus, a job that might ordinarily be regarded as the real work of an editor is distributed in a disaggregated fashion across thousands or millions of individual users. Similarly, the transport service Uber monitors users' locations to measure demand for its services—a task that might otherwise be the work of a professional market researcher.[24] Even where individuals do recognize the value of the work they are doing, they are not well positioned to demand proper compensation for that work because only in aggregate is the labour of value. Since no individual user is pivotal from the firm's perspective, the bargaining power of individuals is minimal. A mechanism could be designed to re-direct revenue generated from exploitation of user-generated content to socially valuable projects, wherein society acts as a proxy for the interests of individual online 'labourers'.

Even where individuals do recognize the value of the work they are doing, they are not well positioned to demand proper compensation for that work. Since no individual user is pivotal from the firm's perspective, the bargaining power of individuals is minimal.

2.1.8 Addressing firm dominance

A common feature of digital businesses is that they scale very well: although the upfront costs of establishing a new service may be large, additional customers can often be served at very low cost. This gives established incumbent firms a cost advantage because they can spread the upfront cost across many existing users. Moreover, incumbents also often benefit from network effects that make their products more attractive than those of smaller rivals. For example, much of the value of Facebook or Skype lies in the large number of users that can be contacted through these platforms—an advantage not enjoyed by smaller rivals, making entry into such markets difficult. These economies of scale on the supply and demand side tend

[21] Martin, K. (2015). Understanding Privacy Online: Development of a Social Contract Approach to Privacy. *Journal of Business Ethics*, 1–19. http://doi.org/10.1007/s10551-015-2565-9.

[22] Etzioni, A. (2000). A Communitarian Perspective on Privacy. *Connecticut Law Review*, 32(3), 897–905.

[23] A distinction can be drawn between such altruistic content creation and scientific initiatives such as crowd-science—see for instance Galaxy Zoo (https://www.galaxyzoo.org)—where the users' labour is used (in scientific research) in a manner consistent with their initial intent. A third category can also be recognised for which the exploitation of labour argument is less relevant: users that generate content for financial benefit on platforms that allow revenue sharing.

[24] Hirson, R. (2015, March 23). Uber: The Big Data Company. Retrieved from http://www.forbes.com/sites/ronhirson/2015/03/23/uber-the-big-data-company/#12c3c52025f4; Frizell, S. (2014, November 19). What Is Uber Really Doing with Your Data? Retrieved from http://time.com/3595025/uber-data/

to promote market structures with a small number of large, dominant firms and many markets in the digital economy indeed appear to be subject to such dominance.

There are well-known harms associated with markets that, rather than being open and competitive, are dominated in this way. Without any effective competition, the dominant firm(s) are able to behave exploitatively, distorting trade and increasing prices in a manner that is good for profits but harmful to overall welfare.[25]

By reason of their tendency to be in such a state of dominance, markets with strong economies of scale (so-called 'natural monopolies') have long been the subject of special treatment in society. Historically, a common approach was to take companies in these markets (such as telephone services or railroads) into public ownership. The more modern approach, though, is to allow privately-owned pseudo-monopolies to operate independently, but subject to conditions that ensure their position of power is not abused to society's detriment. Part of such an arrangement is often regulation, but it may also involve transfers paid by the firm to compensate society for the privileged position it is permitted to enjoy. For example, in several markets governments sell operating licences to incumbent monopolists and the cost of these licenses serves to redistribute some of the value of being a monopolist. To the extent that a market is dominated by a firm less because of its merits and more because of the inherent economics of the product in question, society would appear to be justified in expecting some form of compensation. It may therefore be justifiable to direct a data financing mechanism at 'natural monopolies' in the data economy.

2.2 Pitfalls in data financing

Several potential valid justifications can be found for data financing. However, potential negative effects must also be acknowledged if rational and appropriately restrained real world mechanisms are to be developed. Data financing could hypothetically take the form of a tax at several stages of data collection and processing. Many of the potential downsides of data financing can be linked to distortionary effects of taxation in general.

2.2.1 Distortionary effects of taxation: who bears the tax burden

Unit taxes alter a firm's strategic calculus and therefore have a number of distortionary effects on market outcomes. Firstly, although it may nominally be paid by a firm, a tax that causes the firm to increase its price is, at least partly, passed on to consumers. The extent to which the burden of a tax is passed on to consumers rather than paid out of a firm's profit is known as the tax's incidence. Roughly speaking, the general principle is that the greatest share of the tax burden will fall upon the party that is least price-sensitive. In particular, if a firm's customers would sooner tolerate a price-increase than curtail their consumption of the product then it is likely that they will bear the majority of the burden of a tax. As a consequence, a tax that is intended to capture a share of firms' profits may have the unintended consequence of primarily penalising consumers.

A tax that is intended to capture a share of firms' profits may have the unintended consequence of primarily penalising consumers.

When the product is an online advertisement some caution is needed because the customer is an advertiser rather than a consumer. To the extent that the burden of a tax falls upon these advertisers, one should expect an increase in the prices of the goods that they sell. Thus, the tax burden will ultimately be shared across three parties: the platform, advertisers, and consumers.

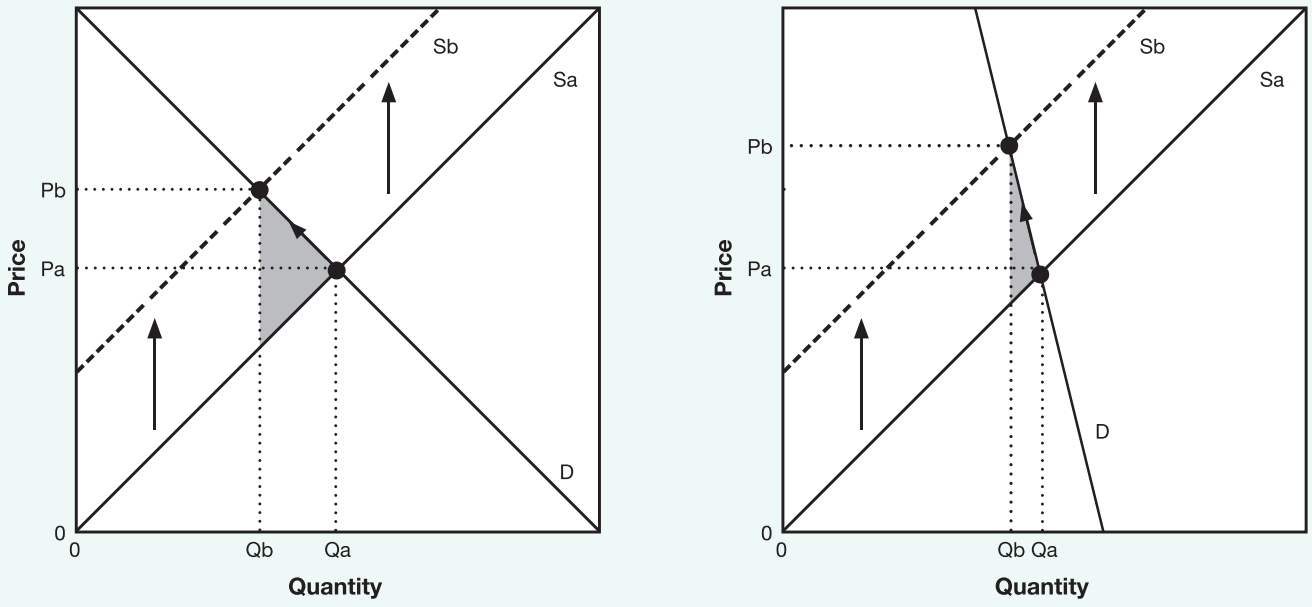
[25] For these purposes, it is possible to treat the personal data that a firm demands of its users as a price. A firm might think twice about demanding too much data if it faces stiff competition, but has more leeway when consumers have no viable alternative to turn to.

2.2.2 Distortionary effects of taxation: foregone trade and welfare loss

Many millions benefit from consuming a wide variety of digital goods and services. A tax on a firm's output discourages it from engaging in such beneficial trades—and the social benefits associated with the lost trade are forgone. The volume of trade lost to taxation, again, depends upon the sensitivity of supply and demand to prices. The distortion will be smallest when demand is highly insensitive to price. A fairly fundamental trade-off is therefore observed: cases where a tax is least likely to reduce trade (demand is price insensitive) are also those in which consumers will bear a disproportionately large share of the tax burden. Conversely, instances where the tax is borne mostly by firms are those in which the social cost of the tax are likely to be greatest. Box 1 provides an illustration of these principles.

The welfare loss associated with a tax distortion grows disproportionately quickly with the size of the tax. It is generally, therefore, more efficient to raise tax revenue from a low rate of taxation applied broadly than a high rate of taxation on a narrow tax base. This has particular relevance to the public goods and positive externalities justifications for data financing identified above. When the main beneficiaries of a public good are data-intensive firms, it may seem natural to propose that those goods should be paid for mostly by those same firms. While such an arrangement would be consistent with norms of social justice, it would typically be less economically efficient than a less targeted mechanism.

Two caveats are in order. Firstly, rather than being harmful, it should be noted that the reduction in trade caused by a tax may be desirable if it corrects for over-consumption associated with a negative externality, such as the overconsumption of attention. As noted



The welfare economics of taxation

These figures show the effects of a tax given two different demand curves, labelled D. The tax increases the minimum price at which firms are willing to supply each unit, causing the industry supply curve to shift upwards from Sa to Sb. This distorts both quantity (falling from qa to qb) and price (increasing from pa to pb). Notice that the magnitude of these changes depends upon how sensitive demand is to price. When demand is price insensitive (the demand curve is very steep), the quantity distortion is small but prices increase quickly– implying that more of the tax burden is borne by consumers. The shaded triangle represents the welfare loss due to taxation, which is larger when demand is more price sensitive. These figures assume the market is highly competitive, but a broadly similar logic applies under other market structures.

above, the presence of such externalities in the data economy could be regarded as a key justification for the introduction of corrective taxation. Secondly, taxes that are applied on a lump-sum basis, rather than in proportion to a firm's sales do not change the profitability of supplying or not supplying each unit, and therefore do not distort the firm's supply decisions in the manner outlined above. The welfare costs noted here therefore do not arise under lump-sum taxation. The main difficulty with lump-sum taxation is in choosing the appropriate level of the tax in a manner that is not perceived as arbitrary or unfair.

2.2.3 Incentives for innovation and investment

Investments in innovation and in infrastructure development yield new products or production techniques that benefit society at large. Some of these benefits accrue to consumers, while a share is retained by the firm in the form of higher profits. It is these profits that constitute the reward for investment and provide the incentive to undertake such investment in the first place. If a firm anticipates that some of the additional profits from its investments will be confiscated via taxation, then its incentive to invest will be reduced. Although there may be a short-term social gain from the taxation regime, the long-run costs from reduced investment could potentially be significant.

The danger of such 'chilling effects' is particularly acute in the case of broadly-conceived levies that are likely to capture as yet unanticipated innovators of the future. More generally, if large, successful firms are targeted for taxation then the prevailing message for entrepreneurs may be that success is likely to be punished. Especially in high-technology markets, where innovation and infrastructure development play an important role, the design of any data financing mechanism should be sympathetic to the need to maintain incentives for investment in these activities.

2.2.4 The availability of alternative instruments

If a data financing mechanism is to be justified on the basis that it is not only a source of revenue, but also a means to achieve some broader social goal then it is important to ask whether there are alternative policy instruments better suited to that end. For example, if the objective is to protect consumers from exploitative data-gathering practices, or to correct for a negative externality, one should check whether the same aims could be achieved more efficiently via privacy regulation or similar interventions. Similarly, if one is concerned with the growing dominance and power of technology companies, then it is relevant to ask whether this is better dealt with through the existing institutions of antitrust policy.

However, data financing mechanisms do not need to be explicitly linked to a perceived failing or opportunity of the data economy. Rather, similar to the UNITAID global health initiative, mechanisms can conceivably be deployed solely to generate revenue for philanthropic or socially beneficial purposes entirely unrelated to the data economy. The justification of a data financing mechanism and the eventual expenditure of capital raised do not need to be connected.

With that said, the case for a data financing mechanism is strongest when it not only achieves some broader social goal, but does so more effectively than any viable alternative approach. While the efficiency of potential mechanisms in these terms is not explicitly considered in this report, it should be part of any future proposal for a real world mechanism.

2.3 Four data financing models

Data financing mechanisms face many additional barriers to implementation beyond the few pitfalls explored here. To explore these, four broad models for a data financing mechanism are now considered.

2.3.1 A global Internet subsidy

One broad class of justifications for data financing concern the need to provide public goods and correct for positive externalities such as network effects in technology adoption through subsidization schemes. For

example, subsidization of Internet access for the poorest members of global society would not only provide access to them, but also benefit other users and online service providers by growing the population of Internet users. More broadly, there are many prerequisites for technology adoption (such as good education) that could be funded under such a scheme. This suggests that a global fund to make strategic investments in development, with a focus on technology and connectivity would be of significant value.

The case for a data financing mechanism is strongest when it not only achieves some broader social goal, but does so more effectively than any viable alternative approach.

Such a fund could be used in several ways to encourage Internet uptake, from funding local start-ups or ISPs to building infrastructure on either proprietary or open technological protocols. A subsidy for global Internet access could be funded in several ways. Online companies could voluntarily contribute to the fund—perhaps according to an agreed metric, e.g. number of users or revenue generated from secondary uses or sales of user data. Alternatively, a levy on Internet users could be justified by the subsidy's reduction of 'digital divides' between online and offline individuals, groups and regions.

2.3.2 A privacy insurance for personal data processing

Processing vast amounts of personal data can be very valuable to business, but it also involves an inherent risk of data leaks. Data leaks can cause harm to data subjects through privacy loss and identity theft as well as to wider society and industry by enabling unauthorised access and ultimately reducing public trust. It could be proposed that responsible data processors should address this possibility pre-emptively, by contributing to a common fund.[26] The fund could be spent on the development of privacy-enhancing technologies, or compensation for victims, thereby providing contributors with a kind of insurance against the costs associated with data loss. Precedent exists for mandating the uptake of insurance where actors take on risks that may affect others—for example, motorists are required to have insurance for third-party damaged and many businesses in some jurisdictions are required to have employer's liability insurance.

Alternatively, in recognition of the harm caused to wider society, some fraction of the revenue from this scheme could be diverted for other social responsibility projects that benefit society, such as the development and public good projects identified above. A suitably judged contribution rate could also serve to address the negative externality inherent in data retention by inducing data controllers to be more circumspect about the data that they hold. The responsible use of such a public insurance scheme might also help data processors to 'insure' themselves against the reputational damaged of a data breach.

2.3.3 A shared-knowledge duty for open and public data

Wealth is not created in isolation and is dependent on economic and political communities. As part of a social contract with data subjects, data processors generating revenue from shared resources may have a duty to share the benefits of processing fairly. There is a case to be made that contributors to a dataset should be entitled to usage of proprietary information, products and services created from processing of public data, without having to share additional personal data or pay the market rate for access, because the creation of the public dataset already constitutes a type of payment. In other words, exploitation of shared resources would need to benefit the public or groups that contribute to the dataset, which can include society in general or specific groups of users conceived as 'digital labourers'. On this basis, a duty could be levied for revenue-generating processing of open and public data. Such

[26] Edwards, L. (2004) Reconstructing Consumer Privacy Protection On-line: A Modest Proposal. International Review of Law, Computers & Technology 18(3): 313-344.

a duty could be applicable to both public datasets that have been funded by public resources or taxation and the development of products or services from the 'free labour' of Internet users (e.g. scraping user content to improve a personalisation algorithm).

2.3.4 An attention levy for digital marketing

A significant application for the processing of personal data has been the online advertising industry. We have argued that users' attention is often exploited and that this may lead both to economic inefficiency and also to violation of individuals' sovereignty over their personal or private sphere. Both issues could be addressed with a corrective levy on intrusive advertising that compels organisations to be more circumspect about invading a user's attention.

2.4 Conclusion

Data financing can be justified by appeal to a broad range of economic, social and ethical goods. From an economic and social perspective, the justifications for data financing are broadly similar to justifications of taxation and philanthropic pursuits. Each mechanism can achieve public goods, create positive externalities and correct negative externalities. Ethical justifications for data financing connect to specific positive and negative externalities. For the former, revenue can be repurposed to fulfil to a greater extent the implicit social contract between data controllers and processors, their users and society more broadly. For the latter, depletion of the attention commons and violations of privacy can be corrected and protected against by mechanisms that encourage responsible advertising, design of digital services and handling of personal data. Two possible economic justifications also address specific negative externalities: the undervaluing of digital 'free labour' involved in the creation of the data that allow the data economy to flourish, and the market dominance of large data-intensive firms.

Each justification offers a potential platform for proposals for future data financing mechanisms. The relative force of each justification is dependent upon the context-specific validity of underlying premises; the existence of a social contract between data processors and users could, for instance, be rejected on the basis that service agreements or consent between social media platforms and users form a prevailing explicit contract. While future proposals will need to be considered within the constraints presented by specific jurisdictions and data processing sectors, the justification of data financing is generally strongest when it achieves a social good more efficiency than alternative viable approaches, such as existing taxation or regulatory regimes.

Each of the four proposed data financing models is based on a social good achievable through data financing. As already noted, data financing mechanisms are most strongly justified when they achieve social goods more effectively than viable alternatives, including existing taxation or regulation. The models are not intended as proposals for implementation, but rather provide a basis for the remainder of the report to identify barriers and opportunities for implementing different types of data financing mechanisms in the future. The next section begins by assessing technical and practical barriers to data financing related to the measurement and valuation of data, and the attribution of liability for data financing to stakeholders in the data economy.

3 TECHNICAL AND PRACTICAL FEASIBILITY

Key Questions

- How can contribution levels for data financing mechanisms be established and monitored?

- Which stakeholders in the digital economy should contribute to data financing?

Summary

- Any data financing mechanism requires a basis for setting the liability for contributions of stakeholders in the data economy. Contributions can be set voluntarily through agreement between stakeholders, or by measuring data or the value of data in some respect.

- Measurement of data and value is not merely a line-drawing exercise: due consideration must also be given to the cost and feasibility of the scheme—which, in turn, depend upon factors such as the observability and auditability of transactions, and the technical requirements for monitoring and administering the financial contributions.

- Data can be easily measured in terms of volume or quantity. Both measures can help assess the latent risks of data collection and storage to data subjects. However, neither measure correlates directly with the effects of data collection and processing in all cases.

- The value of data can also be measured. However, the value of data is non-linear and easily transferrable.

- Data are a heterogeneous and intangible good, so its value is difficult to define. Value can, however, be partially determined by markets that set a price on personal data and records derived from it. Four possible indicators are: financial results, as measured by revenues or net income, divided by the number of personal data records held by the company; market prices for data; economic cost of a data breach; price of data in illegal (cybercrime) markets.

- The value of personal data can also be estimated by individuals. Surveys and economic experiments have been conducted that reveal how much companies would have to pay individuals for their personal information. The results are speculative and lack market validation, but the experimental setting allows for results to be compared across different types of data.

- Many different stakeholders can be argued to have a requirement to pay into data financing schemes. Potential contributors include Internet companies, data-intensive offline companies, digital advertisers, data brokers, networking hardware vendors, cloud service providers, intermediaries including app stores, the Internet Corporation for Assigned Names and Numbers (ICANN), end users.

- While each type of stakeholder can be said to process data, it is not clear that the activities, data types, and value generated by processing across these industries are sufficiently similar to allow overarching data financing mechanisms to be implemented.

Analysis

Assuming data financing is justifiable, a key challenge lies in deciding how contribution levels of different stakeholders in the data economy should be determined. Contributions can be set in many ways. Voluntary schemes could operate as a type of charity, wherein stakeholders self-determine contribution levels. Informal agreement between stakeholders is similarly possible. However, especially if some degree of compulsion is involved, setting a consistent, auditable, and non-arbitrary level for contributions will often require that data be in some way measured or valued. This chapter considers the difficulties of measuring and valuing data as a basis for assigning liability for data financing to stakeholders in the data economy.

3.1 Measuring Data

'Data' is a deceptively simple concept. Data

are often discussed in terms of dichotomies: open or closed; personal or non-personal; research, government or commercial; static or streaming, big or small. A data financing mechanism can in theory be limited based on any of these and similar dichotomies. As commonly understood, the data economy involves collection and processing of data that in some way describes or is produced by humans. Whether such data can be considered personal, meaning they describe an identifiable individual, varies considerably between collection platforms and legal frameworks. For our purposes the basic observation that the data economy involves data about and created by humans is a sufficient starting point.

Data can be measured and valued in numerous ways. Measurement of *volume* (i.e. bits or bytes) is an obvious starting point. Contribution levels can be based on the volume of data processed or stored by a data controller; cloud computing services that charge a set price per megabyte stored or processed show how this principle works in practice. One problem with this approach is that volume and the social effects of processing are not perfectly correlated. Medical images, for example, take up substantially more volume than search records, yet processing the former is not necessarily more privacy invasive than the latter.

'Data' is a deceptively simple concept. Data are often discussed in terms of dichotomies: open or closed; personal or non-personal; research, government or commercial; static or streaming, big or small.

Other measurements face similar challenges. Data can be measured according to the *quantity* of units held or processed (e.g. records, cases, entries, calls), but again a direct correlation between the size of a dataset and its problematic effects is not consistent.

Despite the potential for arbitrariness, direct measurements of data have the advantage of consistency across processing contexts and across some data types. Additionally, financing mechanisms motivated by risks posed to data subjects arguably require some consideration of volume, quantity and similar measures. Consider for example privacy insurance applied to processing of search records. Risks posed by merely collecting and storing data are largely unpredictable because they depend on the meaning that can be generated through novel processing or aggregation.[27] However, while the actual invasiveness of the records depends on how they are used, data storage creates latent risks to data subjects. Although specific risks cannot always be predicted, data subjects nonetheless face risks of future misuse or leakage due to storage. These risks generically increase with the volume or quantity of records stored.

3.2 Measuring Value

The potential arbitrariness of direct measurements of data as a basis for contribution liability can be avoided by instead focusing on the *value* generated from data. Measuring value raises many challenges as well. Firstly, as already mentioned, a multitude of types of data about humans exist (see "What Are Personal Data?"). Secondly, value depends largely on context—the emails addresses of 1,000 CEOs are far more valuable than the emails of 1,000 teenagers for a business-to-business company, but not necessarily for a photo-sharing app.

A broader measurement issue is raised by the fact that, like pharmaceutical patents and copyrights, data are an intangible asset. Intangible property resists measurement and levying because it is easily transferred. A viable data financing mechanism may need to focus on particular types of data storage or transfers rather than data themselves, such as transfers for commercial purposes.

[27] Mittelstadt, B. D., & Floridi, L. (2016). The Ethics of Big Data: Current and Foreseeable Issues in Biomedical Contexts. *Science and Engineering Ethics, 22*(2), 303–341. http://doi.org/10.1007/s11948-015-9652-2

What Are Personal Data?

The OECD Privacy Guidelines define personal data as "any information relating to an identified or identifiable individual (data subject)." This is a broad concept, which includes, by way of example, the following types of personal data:

- User-generated content, including blogs and commentary, photos and videos, etc.
- Activity or behavioural data, including what people search for and look at on the Internet, what people buy online, how much and how they pay, etc.
- Social data, including contacts and friends on social networking sites
- Locational data, including residential addresses, GPS and geo-location (e.g. from cellular mobile phones), IP address, etc.
- Demographic data, including age, gender, race, income, sexual preferences, political affiliation, etc.
- Identifying data of an official nature, including name, financial information and account numbers, health information, national health or social security numbers, police records, etc.

Personal data can be further categorised in various ways. For example, one definition suggests six types of data—covering rudimentary data disclosed to open an account, additional data voluntarily entered by the user, data entrusted to other users, data about a user uploaded by third parties, data about a user's behaviour within the scope of a service, and data that can be indirectly deduced from other observations.,[1,2] More broadly, personal data may be categorised by its use or purpose: Will it be used at the time of collection or stored for later use? Is it collected for first-party use, or on behalf of a third party?

A second way to categorise personal data is to distinguish that which contains personally identifiable information (PII), and that which does not. PII refers to information that can be used to uniquely identify a person, such as a name, address, or social security number. Non-PII includes things such as browsing or search history. But the line is being blurred by edge cases (such as IP addresses) on the one hand, and by new analysis techniques that allow non-PII to be combined in such a way that an individual is identified on the other. These trends have gone some way to eroding the value of the dichotomy between PII and non-PII.

---

[1] Schneier, B. (2009, November 19). A taxonomy of social networking data. Retrieved from https://www.schneier.com/blog archives/2009/11/a_taxonomy_of_s.html.

[2] OECD. (2011). *The role of internet intermediaries in advancing public policy objectives.* Paris: OECD Publishing. doi:10.1787/9789264115644-en.

The value of data is also non-linear. An individual record has little value, but when records accumulate, value is created through synergistic and network effects. By the same token, there may be diminishing returns once sufficiently large volumes are attained. Knowing the volume of data processed or stored reveals little about the value generated. As argued above, size is not a straightforward indicator of quality or risks—high volume does not guarantee better predictions, for example. Rather, value increases when analytic techniques are applied to data that render them meaningful.

3.2.1 Stages of Value Creation

To more thoroughly examine the value of data, it is useful to consider the data lifecycle and the value created at each of its stages. According to a study by the OECD, value is generated from personal data through a series of processes, or a 'value chain' (see: Figure 2). The chain is visualised in four stages: collection and access, storage and aggregation, analysis and distribution, and usage. Many stakeholders are involved at each stage, and some are involved in multiple stages. Each stage might serve as an initial focus point to establish contribution levels based on the value generated from data.

The potential arbitrariness of direct measurements of data as a basis for contribution liability can be avoided by instead focusing on the value generated from data.

The collection of personal data can occur in three main ways: data can be volunteered by users (e.g., when they open a social media account), observed (e.g., through search queries), or inferred based upon analysis and the combination of different datasets (e.g., credit ratings). Predictive analytics can infer invasive personally identifiable information from aggregated datasets that are non-invasive in isolation.[28]

Many Internet companies engage in speculative data collection due to its potential future value, and store a large volume of unstructured data that does not relate to current usage of the service. Such data stores do not directly benefit users (and in fact pose a threat to privacy), yet might increase company valuations. The usefulness (from the controller's perspective) of data stores could serve as another indicator of contributions; for instance, companies holding excessive data based on potential future usefulness could be taxed due to the latent risks storage imposes on data subjects.

The analysis of data is highly correlated with value. Personal data can be combined with other data sources to create detailed personal and consumer profiles. These profiles can be sold in the market to advertisers at a specific price (data exchanges, defined above, provide a platform for such transactions). Again, this transaction offers a possible entry point to assign a value to data.

The end of the personal data value chain is usage, and stakeholders include the users themselves, government and public sector agencies, and businesses. Value here is more difficult to define due to the seemingly limitless variety of uses for data about humans. However, market valuations provide one possible way forward to determine contribution levels.

Personal Data	Collection/Access	Storage and Aggregation	Analysis and Distribution	Usage
Volunteered e.g. declared hobbies and interests, preferences, expertise, etc. **Observed** e.g. location information, browser history, shopping habits etc. **Inferred** e.g. credit ratings, profiles built from online activities, etc.	• Mobile phones • Blogs and discussion lists • Social, professional and special interest networks • User-generated content • Loyalty schemes operated by retailers • Smart appliances • Applications • Sensors etc.	• ISPs and phone providers • Government agencies (e.g. tax offices, property registries, etc.) • On-line social networks • Financial institutions • Medical practitioners • Utility service providers • Retailers etc.	• Retailers and service providers • Public administration • Financial institutions • Healthcare providers • Specialised companies involved in online advertising and market research • Data analysts, providers and brokers etc.	• Business • Government and public sector agencies • End users.

Figure 2 - Personal data value chain

Figure reproduced from: OECD. (2013). Exploring the economics of personal data: A survey of methodologies for measuring monetary value. *OECD Digital Economy Papers,* No. 220. Paris: OECD Publishing. doi:10.1787/5k486qtxldmq-en.

[28] Crawford, K., & Schultz, J. (2014). Big data and due process: Toward a framework to redress predictive privacy harms. *Boston College Law Review, 55*(1). Retrieved from http://bclawreview.org/review/55_1/03_crawford_schultz/

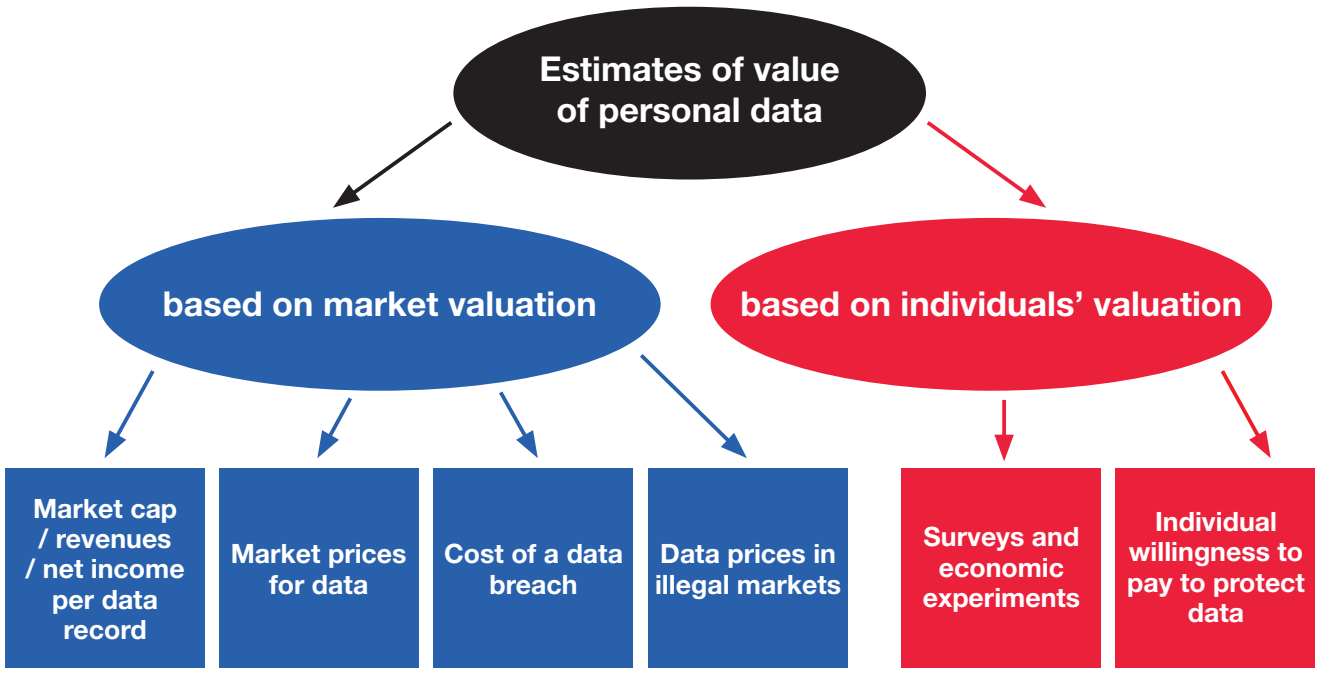


Figure 3 - Estimates of the value of personal data

Figure reproduced from: OECD. (2013). Exploring the economics of personal data: A survey of methodologies for measuring monetary value. *OECD Digital Economy Papers,* No. 220. Paris: OECD Publishing. doi:10.1787/5k486qtxldmq-en.

3.2.2 The Market Valuation of Data

Data are a heterogeneous good, so its value is difficult to define. Value can, however, be partially determined by markets that set a price on personal data and records derived from it. This is one of the main ways in which the monetary value of personal data can be estimated (Figure 3).

A study by the OECD outlined four key indicators based upon market valuation.[29] The first is a simple maths exercise: financial results, as measured by revenues or net income, divided by the number of personal data records held by the company. Directly applied, this exercise has little validity beyond firms that have business models based exclusively on personal data (e.g. social media companies). But, given data on multiple firms in an industry, it may be possible to statistically control for other contributions to revenue estimate the (possibly) non-linear relationship between income and data holdings. Such an estimate would provide an indication of the revenue value of an additional data record. This approach also points to the more general possibility of using information disclosed in financial reports, such as mandatory SEC filings, as bases for calculating contributions to a data financing mechanism. For instance, Amazon's financial reports distinguish the company's cloud computing income from its other sources of income. At best such figures might be construed as rough proxies for the value of some notion of data or data processing, not as direct measures. But a potential advantage of this approach is that financial reports for public companies are readily available and present information in a somewhat standardized form. This could make measurement and compliance significantly more feasible than with any methods based on technical measurements of data flows.

The second indicator is market prices for data: how much is a personal data record sold for on the market? This applies directly to data brokers, but the value is subjective and heavily conditioned by context. A bigger issue is that we cannot see the value of a particular dataset unless there is a transaction in a public marketplace. Transactions involving personal data (e.g., among social media

[29] OECD. (2013). Exploring the economics of personal data: A survey of methodologies for measuring monetary value. *OECD Digital Economy Papers,* No. 220. Paris: OECD Publishing. doi:10.1787/5k486qtxldmq-en.

companies and data brokers), however, are not in public marketplaces. Transparency of transactions is therefore a key condition for this approach to calibrating data financing.

A simple transaction to measure is digital advertising sales. Each time a user requests a page with advertising space, a highly-automated real-time auction is run in the background. In this auction, advertisers are provided with a profile of the user's known attributes and asked to bid for the advertising opportunity. The auction therefore puts a price to user profiles of varying types. However, because the marketplace is not public, only the winners of the bids know what they paid.

More generally, an open and standardized marketplace for business-to-business trade of personal data is theoretically interesting, and applicable to companies with data sharing agreements. In practice, though, such a market has not emerged because very few data companies are likely to want to publicly sell their proprietary data. It is in their commercial interest to sell privately so that the price can be adjusted for different customers. Data are not homogeneous, and value depends on the customer and context.

The third indicator is the economic cost of a data breach: how much do companies have to pay when a data breach involving personal data occurs? This equates value with risk incurred by data storage, and is related to the privacy insurance model. This indicator however captures costs related to damage caused as opposed to the data involved.

Data are not homogeneous, and value depends on the customer and context.

The fourth indicator is the price of data in illegal (cybercrime) markets, which are used to exchange personal data. Credit card information is commonly stolen and resold, as well as usernames and passwords. The main problem with this approach is that it values data in illegal use, and is therefore of limited utility for estimating the legitimate economic value of data. For example, a criminal may pay $100 for a bundle of valid credit card numbers in anticipation of using them to benefit from several thousand dollars' worth of fraud. This bears no relation to the value of a credit card number to organizations that are constrained to use it only for legitimate means.

The value of personal data can also be estimated by individual data subjects (Figure 3). Surveys and economic experiments have been conducted that reveal how much companies would have to pay individuals for their personal information. The results are speculative, potential subject to behavioural biases and lack market validation, but the experimental setting allows for results to be compared across different types of data.

3.3 Data Economy Stakeholders

Beyond types of data and measurements, many different stakeholders can be argued to have an obligation to pay into data financing schemes. The breadth of firms that process, control, or otherwise benefit from data about humans is daunting.

Data financing mechanisms could be applied to digital adverting, streaming services, free exchanges, search engines, databases, loyalty card schemes, credit rating agencies, financial institutions, and intermediary platforms (such as app stores), to name only a few possibilities. Each have business models that are premised upon the collection and sales of personal data. While each can be said to process data about humans, it is not yet clear that the activities, data types, and value generated by processing across these industries are sufficiently similar to allow for overarching financing mechanisms.

The breadth of firms that process, control, or otherwise benefit from data about humans is daunting.

3.3.1 Internet Companies

Many companies conduct business exclusively or primarily on the Internet, and have achieved substantial market valuations as a result of business models that are driven by vast volumes of personal data. Three categories of Internet companies are

I Want to Pay for Facebook

Advertising-financed services like Facebook are free only in the narrow pecuniary sense. Their currencies are privacy and attention. How much exactly does Facebook earn from our data? Not very much, as it turns out: Zeynep Tufekci estimates that Facebook makes twenty cents per-user per-year.[3] The drive to achieve even this level of per-user profit by improving customer targeting has, though, distorted the content and design of Facebook and other ad-based platforms. Every facet of site design—from the algorithms that promote content to prominence through to decisions about layout and navigation—is informed, in part, by optimisation of the advertising operation. In the words of Tim Wu, "To pay with data is to make yourself more vulnerable to the outside world... The more data you give away, the more commercially customized your world becomes. It becomes harder to ignore advertisements or intrusions."[4]

Tufekci and others have proposed a simple solution: allow users to pay for services directly, opting-out of the advertising funded model. This would make explicit to users the economic nature of the transaction and offer some protection from the distortions described above. However, the proposal faces a number of important challenges. Firstly, there is an adverse selection problem: those most likely to opt-out are the wealthiest users who are also most valuable to advertisers. This implies that the cost of opting-out might have to be significantly above the level implied by the average value of a user's data. Secondly, the two-sided platform business at the centre of online advertising has proven to be a formidable and fairly sustainable source of revenue that the industry is likely to be reluctant to abandon.

[3] Tufekci, Z. (2015, June 4). Mark Zuckerberg, let me pay for Facebook. *The New York Times.* Retrieved from http://www.nytimes.com/2015/06/04/opinion/zeynep-tufekci-markzuckerberg-let-me-pay-for-facebook.html?_r=0.

[4] Wu, T. (2015, August 14). Facebook should pay all of us. *The New Yorker.* Retrieved from http://www.newyorker.com/business/currency/facebook-should-pay-all-of-us.

particularly salient: online retailers like Amazon, which collect purchasing data; social media platforms like Facebook and Twitter, which collect demographic data, social data, and user-generated content; and search engines like Google, which collect search query data. All of the companies collect general activity or behavioural data generated by clicks on a webpage, which reveal greater insights about who their users and customers are. While measuring the monetary value of personal data can be difficult, it is undeniable that the data are considered valuable by the company collecting them.

3.3.2 Data-Intensive Offline Companies

Although tech companies have been quick to embrace the opportunities provided by an explosion of data, many traditional firms have similarly acted upon this opportunity. Brick and mortar retailers, for example, offer loyalty card schemes.

3.3.3 Digital Advertisers

Broadly speaking, there are three players in the online advertising industry that finances the Internet: advertisers (companies that wish to promote their products and services), web publishers (companies that display ads on their websites), and ad networks. Ad networks occupy a central position, as they are intermediaries that purchase ad space from publishers and resell to advertisers. Search engines, media companies, and technology vendors are all involved in this space.

Digital advertisers are an obvious candidate for any type of attention levy or privacy insurance, as ads attempt to capture the attention of users, often through explicit or inferred knowledge about the user's behaviours and preferences. Ad networks employ three strategies to match advertisers with customer segments: contextual (e.g., keywords), vertical (publishers in a similar industry are grouped), and behavioural (browsing behaviour is used to categorise

interests). Behavioural advertising based on user tracking is particularly relevant in terms of user privacy. Potentially invasive advertising is made possible by the "tracking ecosystem", in which all of a user's clicks are recorded and sold to advertisers. Various tracking techniques identify unique characteristics of users' computers, which allow advertisers to target the users themselves.[30]

Data exchanges—marketplaces "where advertisers bid for access to data about customers"—can be used as a nexus for data financing mechanisms.[31] Transactions in these markets provide an efficient common proxy measure of 'data' for compulsory data financing mechanisms.

3.3.4 Data Brokers

Sometimes called information resellers, data brokers are "companies that collect and aggregate consumer information from a wide range of sources to create detailed profiles of individuals".[32] Sources include public records, social media, and self-reported data. This information is sold to other companies and organisations, and the user is not involved in the transaction.

Some data brokers contribute to risk mitigation by assisting clients with identity verification and fraud detection.[33] Others, such as Acxiom, are marketing-oriented and help clients craft targeted marketing messages to sell more products or services. It is this category of data brokers that might be subject to a data financing contribution. Transparency of transactions can, however, prove a barrier to the involvement of data brokers. The U.S. Federal Trade Commission has published a report that advocates greater transparency and accountability in the data broker industry, recommending the introduction of legislation that requires data brokers to provide consumers with access to their data, as well as the ability to opt out of it being shared for marketing purposes.[34]

3.3.5 Networking Hardware Vendors

Networking hardware devices—equipment that capture the flow of data across the network—can provide a proxy for data processing for the sake of apportioning liability for contributions to a financing mechanism. These devices process data in an indiscriminate manner (there is no focus on personal data), and are manufactured by companies like Cisco, Netgear, and Ericsson. Implementing a data financing mechanism on such devices would be akin to the private copying levy on recordable media, which is used to indirectly compensate artists and performers (copyright owners) for losses due to piracy. This levy has met with resistance in many countries, as recordable media is also used for content that is not copyrighted.[35] Similar challenges can be raised for a proposed financing mechanism on networking hardware, which generically facilitate data transfer and so are not limited to types of data for which financing mechanisms are particularly justifiable.

3.3.6 Cloud Service Providers

Data storage has migrated from hardware to 'the cloud'. Cloud service providers (CSPs) such as IBM and Microsoft are helping many information technology companies store their data. Cloud services providers already (normally) price their services based on volume of data stored or transferred. Service providing can be a consistent and

[30] Nikiforakis, N., & Acar, G. (2014, July 25). Browser fingerprinting and the online-tracking arms race. *IEEE Spectrum*. Retrieved from http://spectrum.ieee.org/computing/software/browser-fingerprinting-and-the-onlinetracking-arms-race.

[31] eXelate. (2016). *Data marketplace.* Retrieved August 4, 2016, from http://exelate.com/products/data-marketplace/.

[32] Privacy Rights Clearinghouse. (2016, May). *Fact sheet 41: Data brokers and your privacy.* Retrieved August 4, 2016, from https://www.privacyrights.org/content/data-brokers-and-your-privacy.

[33] Ibid.

[34] Federal Trade Commission. (2014, May). *Data brokers: A call for transparency and accountability.* Retrieved August 4, 2016, from https://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf.

[35] ARIA. (2016). *The blank media levy—Not in the interests of artists or record companies.* Retrieved August 4, 2016, from http://www.aria.com.au/pages/TheBlankMediaLevy-NotInTheInterestsofArtistsorRecordCompanies.htm.

easily measurable proxy for data to apportion liability for contributions to data financing.

3.3.7 App Stores

Apple and Google dominate this domain, which represents a chokepoint for businesses looking to learn more about users through mobile data. In 2015, the iOS App Store generated over $20 billion in gross revenue, while Google Play generated over $10 billion.[36]

Users generate vast volumes of personal data through apps; real-time location data are particularly informative. App permissions (e.g., access to address book, location tracking) reflect the level of personal data requested by a given app, and thus provide a potential proxy for data financing. A financing mechanism could conceivably use the number of downloads and permissions required by an app to apportion liability for contributions to the app store or developer.

3.3.8 The Internet Corporation for Assigned Names and Numbers

The Internet Corporation for Assigned Names and Numbers (ICANN) represents another 'data chokepoint' where a financing mechanism could be embedded. ICANN manages the domain name system, and receives revenue every time a domain name is purchased. While valuable for its ubiquity, domain names bear no relevance to revenue or volumes of data processed.

Mechanisms for social good have already been proposed for ICANN. The European Commission (EC) has previously argued for greater governance of ICANN. The EC has suggested that an 'independent control mechanism' be established to monitor ICANN's finances, and that surplus from registry operations be transferred to 'an appropriate public interest beneficiary' that addresses, for example, development issues. This call aligns well with the aims of data financing.[37]

Moreover, the generic Top-Level Domain (gTLD) programme has significantly increased ICANN's revenue. In March 2016, the one thousandth gTLD was sold.[38] Companies have been purchasing not only their own names, but also generic nouns like 'book' and 'bar'. Such nouns are arguably a depletable natural resource. Selling a gTLD to a private interest is against the spirit of trademark law, because generic nouns cannot be trademarked. Thus, a case can especially be made for a slice of the proceeds from the sale of gTLDs to contribute to development projects.

3.3.9 End Users

As detailed above (see: Section 2.2.1), end users may bear the cost of data finance mechanisms if affected data intensive firms shift the costs downstream. However, mechanisms can also explicitly target end users. Taxing end users may also be unjustified due to the balance of power and information asymmetry that favour data processors in personal data exchanges. Individuals are not well placed to understand the potential value of personal data, due both to its scale and the complexity of methods to generate meaning and value from it.[39] However, comparable to levies on carrier bags that reduce consumption,[40] data financing mechanisms can aim to raise awareness of the potential value of personal data among end users. Data subjects may demand greater access, control or a share in revenue generated from personal data processing as a result.

Comparable mechanisms have already met with resistance. In Hungary, a proposed Internet usage levy was perceived as deeply undemocratic.[41]

[36] Woods, B. (2016, January 20). Google Play had twice as many app downloads as Apple's App Store in 2015. *The Next Web.* Retrieved from http://thenextweb.com/apps/2016/01/20/google-play-had-twice-as-many-app-downloads-as-apples-app-store-in-2015/#gref.

[37] European Commission. (2011, September 1). *EC paper on ICANN: Finances.* Retrieved from https://www.evernote.com/shard/s1/share/4846-s1-b64562924a5c1628c.

[38] ICANN. (2015). *FY15 adopted operating plan and budget.* Retrieved from https://www.icann.org/en/system/files/files/adopted-opplan-budget-fy15-01dec14-en.pdf.

[39] Tene, O., & Polonetsky, J. (2013). Big data for all: Privacy and user control in the age of analytics. Retrieved from http://heinonlinebackup.com/hol-cgi-bin/get_pdf.cgi?handle=hein.journals/nwteintp11&section=20

3.4 Conclusion

It is clear that data financing mechanisms raise multiple challenges for determining and monitoring contribution levels for different stakeholders. For compulsory data financing schemes, measurement is a critical issue because data are an intangible and highly varied asset. Volume and quantity can both be used, but neither provide a universally applicable rational and non-arbitrary basis for contributions to a data financing mechanism. Both are heavily dependent on the context, stakeholders and data types involved. Measurements of the value of data are an alternative. Markets and individuals can set prices on data that are specific to sectors and data types where a data financing mechanism can be implemented. However, as the value of data is largely determined by its potential meaning, prices set by markets cannot possibly represent permanent value. Voluntary schemes can avoid many of these problems if contributions are based on agreement between contributing stakeholders and governance bodies. Cooperation may not however be straightforward given the variety of potential stakeholders that can contribute to data financing, and competition between them.

[40] Ritch, E., Brennan, C., & MacLeod, C. (2009). Plastic bag politics: modifying consumer behaviour for sustainable development. *International Journal of Consumer Studies, 33*(2), 168–174. http://doi.org/10.1111/j.1470-6431.2009.00749.x

[41] Euronews. (2014, October 29). "We don't pay tax to criminals": Hungarian anger at Internet levy. Retrieved from http://www.euronews.com/2014/10/29/we-dont-pay-tax-to-criminals-hungarian-anger-at-Internet-levy.

4 POLITICAL AND LEGAL FEASIBILITY

Key questions

- What are the key political requirements facing data financing?

- What legal challenges will face data financing internationally?

Summary

- There are four possible scenarios for implementing a novel financing mechanism: a unilateral or multilateral CSR scheme or, alternatively, a mandatory data financing mechanism introduced at either the national or supranational level. Each of these scenarios faces political and legal challenges.

- There is clear political momentum for introducing measures that would be seen as promoting compliance with data ethics, cybersecurity and data protection regulations. Both national governments and businesses are increasingly keen to address issues surrounding safety on the Internet and the ethics of large-scale data.

- Ideally, implementation of a financing mechanism linked to data should be international, as digital data transcends national borders. However, any attempts at international legal implementation could stumble upon conflicting political agendas and national legal regimes.

- Given the difficulties associated with unilateral implementation of mandatory legal regimes, the importance of international cooperation cannot be underestimated. Supranational measures, however, require reconciliation of conflicting political agendas, which is never an easy task.

- A supranational tax solution would be challenging to implement. A key obstacle is to ensure that national governments agree on a common set of rules regarding calculation of contributions, which ultimately determines how to share the burden imposed by a data financing scheme.

- A novel financing measure could also be expected to require some sort of international governance regime, which would administer collection and distribution of funds.

- The legal foundation for the implementation of data protection measures exists, both at the EU and the UN levels. However, differences in legal regimes across key players, namely the US, the EU and China, might hamper any further attempts at a truly global implementation of the regime.

- A key concern in designing a data financing mechanism is the existence of more appropriate legal mechanisms, such as corporate tax regimes, antitrust laws and personal data protection regulations. Any novel financing measure introduced by way of a legal instrument would have to meet certainty requirements, which might be challenging to achieve given the definitional problems relating to data.

- The most plausible scenario, given political momentum, would be to pursue CSR initiatives at unilateral and multilateral levels, while investigating possible new governance regimes that could be put in charge of administering the novel financing mechanisms.

Analysis

Assuming appropriate stakeholders can be identified and value assigned to data or a proxy thereof, political and legal challenges still remain for data financing. As with the Internet, data financing is inherently trans-national. Data flows across legal and national boundaries create jurisdictional issues and conflicts between legal frameworks. This chapter identifies key political and legal barriers to data financing at national and international levels.

Two critical dimensions of any proposed financing mechanism are the degrees of compulsion and multilateralism. Possible mechanisms can range from

entirely voluntary charitable contributions, comparable to existing CSR initiatives, to compulsory taxes. Similarly, mechanisms can range from decentralised schemes administered by individual organisations to multilateral cooperative efforts. Multilateral partnerships allow for broader scale data financing mechanisms than decentralised schemes in principle, while requiring stronger governance and political cohesion. Meanwhile, while ideally voluntary schemes alone would provide sufficient funding to meet global development needs and the fulfilment of unmet social needs related to the data economy (see: Chapter 2), compulsory mechanisms may nevertheless be required to mobilize sufficient capital.

Two critical dimensions of any proposed financing mechanism are the degrees of compulsion and multilateralism.

Figure 4 includes examples of existing finance mechanisms both within and beyond the data economy. As demonstrated throughout this chapter, these dimensions shape the political and legal barriers facing data financing mechanisms.

4.1 Political requirements

As with any policy initiative, novel data financing mechanisms require political support to be implemented. Policy-makers and industry must be sufficiently willing or compelled to implement and adhere to a mechanism.

Political momentum, international cooperation and authoritative governance arrangements are key requirements for future data financing mechanisms.

4.1.1 Political momentum

Despite its novelty, momentum in related governance and regulatory arenas suggests that support may be available for data financing if it can be linked to corporate social responsibility. Digitalisation of business processes has been widespread for many years, but recently executives have started implementing comprehensive digital strategies for addressing cybersecurity risks

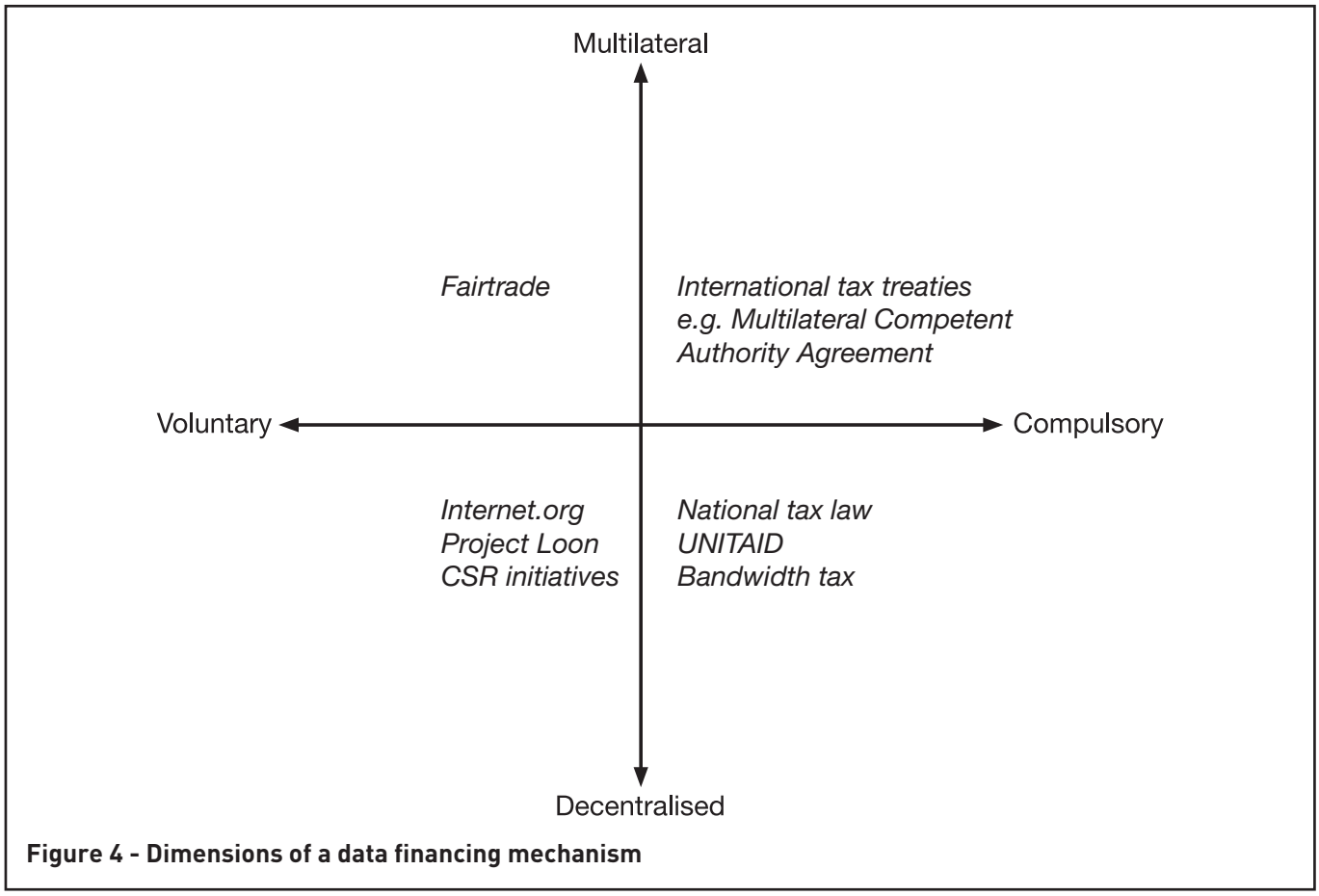


Figure 4 - Dimensions of a data financing mechanism

as part of the overall drive towards better corporate digital governance. From 2012 onwards the advent of corporate big data, the growing importance ascribed to data science methodologies, and the attention given surrounding machine learning and artificial intelligence have all brought far more widespread corporate concern for the ethics of data than we have seen before. There is, therefore, significant momentum to set up compliance and ethics committees within corporate boards in order to oversee practical implementation of digital security and data protection regulations.[42]

Policy-makers and industry must be sufficiently willing or compelled to implement and adhere to a mechanism.

A financial contribution, such as a CSR donation, could be seen as a contribution towards better compliance with data protection and cybersecurity. This might be attractive to corporate executives, who would be keen to be seen to be improving corporate digital governance. However, a chief obstacle to acceptance of novel financing initiatives would be the fact that digital multinationals already contribute to CSR projects, such as Internet infrastructure investment as part of Google's Project Loon and Facebook's Internet.org.[43] Given the level of current investment, a more plausible scenario would need to look at giving innovators tax incentives for making infrastructure investments.

In the government arena, policy officials are also keen to ensure that sufficient regulatory mechanisms are implemented to protect the business community and the general public from the security threats that increased digitalisation entails. Greater political momentum and willingness to act can be seen among EU officials as they adopt detailed industry codes of conduct, particularly for health apps, that would regulate excessive collection and storage of personal data and issues around users' consent.[44] Furthermore, a new EU security standards directive will introduce a robust framework for digital service providers, such as online marketplaces, cloud providers and search engines, in order to ensure that they adopt a corporate cybersecurity strategy, implement appropriate security measures and comply with notification requirements whenever data leaks occur.[45]

A number of other developments are leading to greater digitalisation and heavier data flows, all of which elevate these issues in the political climate. Further digitisation of

Project Loon: Internet Without the Infrastructure

Project Loon is Google's ambitious attempt to provide high-speed cellular internet to rural areas via a network of balloons in the stratosphere.[5] Pioneered in June 2013 in New Zealand, the project was subsequently tested in Brazil, California, Australia, Indonesia, and most recently Sri Lanka. Through partnerships with telecommunications companies, Google has enabled individuals to connect to the balloon network from their phones. When the testing phase is complete and a ring of balloons is launched into the stratosphere, commercialisation will be possible—wireless carriers will be able to rent balloons to enlarge their networks. The balloon network is an attractive option for developing countries, as it is far cheaper than the installation of underground fibre cables or cell towers.

[5] Google. (n.d.). *Project Loon.* Retrieved August 4, 2016, from https://www.google.com/loon.

[42] EY. (n.d.). Why Digital Governance Matters. Retrieved August 4, 2016, from http://www.ey.com/GL/en/Services/Advisory/EY-why-digital-governance-matters; Richards, N. M., & King, J. H. (2014). Big Data Ethics. *Wake Forest Law Review, 49,* 393.

[43] Project Loon. (n.d.). Retrieved August 4, 2016, from https://www.solveforx.com/loon/; Internet.org by Facebook. (n.d.). Retrieved August 4, 2016, from https://info.Internet.org/en/.

[44] European Commission. (2016, June 7). Code of Conduct on Privacy for mHealth Apps Has Been Finalised. Retrieved August 4, 2016, from https://ec.europa.eu/digital-single-market/en/news/code-conduct-privacy-mhealth-apps-has-been-finalised.

[45] Directive (EU) 2016/1148 of the European Parliament and of the Council of 6 July 2016 concerning measures for a high common level of security of network and information systems across the Union. Retrieved from http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016L1148&from=EN.

The Global e-Sustainability Initiative

The Global e-Sustainability Initiative (GeSI) provides an example of a multilateral collaboration for positive change in the telecommunications industry. GeSI promotes social and environmental sustainability; their objective is to build "a sustainable world through responsible, ICT-enabled transformation."[6] They disseminate unbiased information, resources, and best practices, create tools, and contribute to relevant policies regarding sustainability. Forty of the biggest ICT companies are members of GeSI, including AT&T, BT, Fujitsu, Microsoft, Nokia, and Verizon. GeSI also partners with over twelve organisations—including the Carbon Trust, the United Nations Environment Programme, and the World Resources Forum—on various sustainability projects. GeSI demonstrates that leading ICT companies are able to collaborate with competitors, civil society, and academia at a global level if their interests can be protected.

[6] GeSI. (2016). *Global e-Sustainability Initiative.* Retrieved August 4, 2016, from http://gesi.org.

the global economy could spur the drive to comprehensive adoption of digital governance initiatives. Widespread adoption of the Internet of Things is likely to lead to larger volumes of data flows, thus creating greater exposure to risks of data leaks. Similarly, the anticipated move towards adoption of electronic contracts using blockchain technology could boost the volume of data-intensive transactions, which would necessitate even more stringent security measures that require extra financing.[46]

4.1.2 Challenges of Implementation

The quickest way to implement a financing mechanism would be for states to lead by example and hence push for unilateral measures, either at the national or the company level. However, in practice this can be challenging. A number of state authorities have already failed to tackle the data-intensive industries unilaterally using different iterations of the so-called 'Google tax'. Italy has abandoned plans to require that only Italian resident companies could buy online advertising from foreign companies.[47] Germany's unilateral attempt to introduce a copyright levy on news aggregators has resulted in Google simply exiting the market by shutting down German residents' access to its news service.[48]

Given the difficulties associated with unilateral implementation of mandatory legal regimes, the importance of international cooperation cannot be underestimated. Supranational measures, however, require reconciliation of conflicting political agendas, which is never an easy task. While the results of the Brexit referendum do offer opportunities for redesigning the national legal regime in the UK, it is unlikely that EU law could be completely ignored if access to the single market were still on the agenda. Equally, the EU might be keen to support its digital businesses' competitiveness by investing heavily in the digital single market initiatives. However, it is worth bearing in mind that the biggest players in the digital industry are US corporations, meaning that EU initiatives could come into conflict with US policy.

Even if there is sufficient political momentum, as the regulation of the financial transaction tax at the EU level has demonstrated,

[46] Government Office for Science. (2016). *Distributed Ledger Technology: Beyond Block Chain. A report by the UK Government Chief Scientific Adviser.* Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/492972/gs-16-1-distributed-ledger-technology.pdf.

[47] Italy Cancels Google Tax on Web Companies. (2014, February 28). Retrieved August 4, 2016, from http://uk.reuters.com/article/us-italy-tax-idUKBREA1R0WM20140228.

[48] Meyer, D. (2016, March 24). EU Lawmakers Are Still Considering This Failed Copyright Idea. Retrieved from http://fortune.com/2016/03/24/eu-ancillary-copyright/

[49] Brunsden, J. (2016, June 5). EU financial transaction tax progress stalls. *Financial Times.* Retrieved from http://www.ft.com/cms/s/0/ab4ad04c-29ae-11e6-8ba3-cdd781d02d89.html#axzz4GKzTewap; Dizard, J. (2016, July 29). Democrats Resurrect "Insane" Financial Transaction Tax. *Financial Times.* Retrieved from http://www.ft.com/cms/s/0/fc718e7c-5571-11e6-9664-e0bdc13c3bef.html#axzz4GKzTewap.

a supranational tax solution would be challenging to implement.[49] The key obstacle is to ensure that national governments agree on a common set of rules regarding calculation of contributions, which ultimately determines how to share the burden imposed by a data financing scheme. It is unrealistic to expect that a select few nations would be able to come up with a solution quickly.

Given the difficulties associated with unilateral implementation of mandatory legal regimes, the importance of international cooperation cannot be underestimated.

4.1.3 International governance

Assuming sufficient political momentum exists to launch a data financing mechanism, governance poses a further barrier. Any governance structure must be sufficiently authoritative and representative of the jurisdictions involved to administer collection and distribution of funds. Consensus between states on the design of such a governance structure within the framework of existing international mechanisms may prove difficult. As controversies surrounding Internet governance demonstrate, pre-existing global institutions may be entangled in disputes and historical conflicts. The debate around control of the DNS root by ICANN provides a perfect illustration of the difficulties that might arise.[50]

It may therefore be necessary to set up entirely new governance regimes that would be in charge of implementing data financing mechanisms. The Medicines Patent Pool (MPP) is a good illustration of how such a regime could be implemented in practice.[51] MPP is a United-Nations-backed public health organisation, which is funded primarily by UNITAID. MPP administers a pool of medicine patents, which pharmaceutical companies contribute in order to help manufacture generic drugs to treat such diseases as HIV, hepatitis C and tuberculosis. MPP, in a sense, acts as a one-stop-shop intermediary between inventors and drug manufacturers, helping to increase availability of medicines while guaranteeing a fair royalty stream for innovators. Following a similar logic, a new governance body could be established that would administer novel data-linked financing mechanisms. For example, one of the options could entail setting up a public personal data exchange that seeks a high level of compliance with privacy and security standards when trading in personal data.

The issue, however, is the legitimacy of such governance mechanisms. The need for political momentum and a feasible implementation plan while retaining legitimacy implies a need for trans-national political and industrial cooperation. Non-governmental organisations can administer CSR schemes with mere agreement between data-intensive firms, whereas mandatory legal regimes require multilateral agreement at a national level and harmonisation across differing legal frameworks. Further, any governance body would be entangled in conflicts of interest if funds raised as part of the data financing scheme were subsequently distributed to fund projects from which digital businesses would benefit directly. It might therefore be necessary to divert the funds to other initiatives, unrelated to the Internet and data industries, which might, however, undermine the credibility of any such scheme. Labelling the mechanism as a data financing mechanism would help to explain its purpose.

[50] Twist, J. (2005, November 18). Controversy Blights UN Net Summit. *BBC*. Retrieved from http://news.bbc.co.uk/1/hi/technology/4450474.stm; Pohle, J. (2015, January 5). Multistakeholderism Unmasked: How the NetMundial Initiative Shifts Battlegrounds in Internet Governance. Retrieved from http://www.globalpolicyjournal.com/blog/05/01/2015/multistakeholderism-unmasked-how-netmundial-initiative-shifts-battlegrounds-Internet; BBC News. (2012, December 10). Divisions Over Internet Governance Intensify in Dubai. Retrieved August 4, 2016, from http://www.bbc.co.uk/news/technology-20661932.

[51] Medicines Patent Pool. (n.d.). Retrieved August 4, 2016, from http://www.medicinespatentpool.org/.

It may therefore be necessary to set up entirely new governance regimes that would be in charge of implementing data financing mechanisms.

4.2 Legal challenges

Legal challenges for data financing will vary considerably within complex international, national and regional regimes. In this section initial barriers are considered in two of the world's largest data markets: The European Union and the United States. However, data financing mechanisms proposed in the future must take into consideration the interaction between the legal issues highlighted here, and those emerging within the proposed jurisdictions.

4.2.1 International legal challenges

As already discussed, international implementations of data financing would best reflect the non-territorial nature of data. A solid international foundation for building a legal regime aimed at protecting human rights in the context of business activities does exist. The Ruggie Principles, endorsed by the United Nations and adopted by multinationals, require states and businesses to undertake steps to ensure respect for human rights in the course of business operations.[52] At the European level, the ECHR and, within the EU, the TFEU and the EU Charter of Fundamental Rights, among others, afford protection to privacy and personal data.[53] These legal foundations could become a stepping stone in developing novel financing mechanisms for privacy protection.

Despite the existing international foundations, data financing mechanisms might still be challenging to implement at an international level given the differences between national legal regimes that define and protect privacy and personal data. While the EU will soon require digital multinationals to comply with an all-encompassing regulation (the General Data Protection Regulation; GDPR), the US has a web of complex, but nevertheless patchwork, privacy laws that do not reconcile with the EU regime.[54] To overcome this gap, the newly implemented EU-U.S. Privacy Shield aims to provide a legal basis for data sharing.[55] Nevertheless, tensions remain between European and American legal frameworks for privacy and data protection. Furthermore, the 1998 US Internet Freedom Tax Act bans any tax on Internet access, and various other Internet-specific taxes, such as taxes on bandwidth, bits and email. It also bans multiple taxes on electronic commerce. The law recently became permanent when President Obama signed the Trade Facilitation and Trade Enforcement Act of 2015 (Pub.L. 114–125) on February 24, 2016.

National implementations outside the EU, however, may still be possible, for example, in China or India. The future exit of the UK from the EU could also offer opportunities for designing a data protection regime in the UK that could incorporate elements of novel financing mechanisms aimed at protecting privacy. Given the difficulties with mandatory implementation, however, a CSR exercise looks more realistic in the short term.

Despite the existing international foundations, data financing mechanisms might still be challenging to implement at an international level given the differences between national legal regimes that define and protect privacy and personal data.

[52] United Nations. (2011). *Guiding Principles on Business and Human Rights.* Retrieved from http://www.ohchr.org/Documents/Publications/GuidingPrinciplesBusinessHR_EN.pdf.

[53] European Parliament. (2016, June). Personal Data Protection. EU Fact Sheets. Retrieved August 4, 2016, from http://www.europarl.europa.eu/atyourservice/en/displayFtu.html?ftuId=FTU_5.12.8.html.

[54] Weiss, M. A., & Archick, K. (2016). US-EU *Data Privacy: From Safe Harbor to Privacy Shield.* Congressional Research Service. Retrieved from https://www.fas.org/sgp/crs/misc/R44257.pdf.

[55] European Commission. (2016). Commission Implementing Decision of 12.7.2016 pursuant to Directive 95/46/EC of the European Parliament and of the Council on the adequacy of the protection provided by the EU - U.S. Privacy Shield (No. C(2016) 4176 final) (p. 44). Retrieved from http://ec.europa.eu/justice/data-protection/files/privacy-shield-adequacy-decision_en.pdf

4.2.2 Cutting across existing mechanisms

When it comes to using data financing as a way to provide insurance against data leaks, the key obstacle is the availability of private corporate insurance, which covers privacy breach. A 'privacy insurance' mechanism would be in competition with private insurance companies. Even if such a route could be followed, the high risk of privacy breach associated with digital business might result in a having to set a prohibitively high price for mandatory insurance. However, other mechanisms may diminish this concern: a duty of care, for example, exists in case law around cybersecurity to solve disputes involving negligence.[56]

Data financing may also overlap with regulatory mechanisms better placed or equipped to enhance the social goods sought. In the case of privacy insurance, robust regulatory mechanisms already exist that appear to be more appropriate for data protection. The aforementioned GDPR appears initially better equipped for personal data protection, as it contains legally enforceable mechanisms, including fines and detailed enforcement procedures for data misuse. Proposed mechanisms will ideally target social needs insufficiently met by current regulatory mechanisms.

Finally, it would be difficult to attach a value to privacy rights in order to protect data security. The non-territoriality of data renders ownership rights of the businesses that collect and store personal data tenuous. There is also a conflict between personal rights over collected data and proprietary rights that businesses have over their investment in the collection and processing of data. Commodification is not part of the classic bundle of proprietary rights, so there are arguments against treating personal data as capable of being owned by someone else in the first place.[57]

4.2.3 Legal certainty and definitional ambiguity

From a legal perspective, the implementation of a novel data financing mechanism would be challenging because of numerous definitional ambiguities. As highlighted in Chapter 3, one of the key concerns is how to legally define data to ensure certainty of liability for contributions. Difficulties might ensue when distinguishing between open, private and hybrid data and when compiling a list of exemptions for education, press, research, law enforcement and similar data collection and processing purposes. A vital distinction would need to be drawn between personal and non-personal components of data flows, as transfers of non-personal data are actually encouraged under the current EU policy for cloud computing and geolocation.[58] Any financing mechanism that discourages such non-personal data flows could conceivably be considered a measure that hinders the free flow of goods and services, and would thus violate a central principle of the EU single market. Corporate payers would also be difficult to define without discriminating against either small and medium-sized enterprises (SME) or multinationals; in this regard, hitting businesses with extra levies could contradict Article 16 of the EU Charter of Fundamental Rights concerning the freedom to conduct business in the EU.

4.3 Conclusion

There is some political momentum for establishing a data financing regime that would help to ensure Internet development and data protection. Growing concern over the ethics of big data, data analytics and machine learning may create a window of opportunity for data financing mechanisms. There are, however, a number of significant political and legal barriers that would need to be overcome before any financing regime could be introduced. The key obstacles are the lack of legitimate international governance regimes,

[56] Shackelford, S. J., Proia, A. A., Martell, B., & Craig, A. N. (2015). Toward a Global Cybersecurity Standard of Care: Exploring the Implications of the 2014 NIST Cybersecurity Framework on Shaping Reasonable National and International Cybersecurity Practices. *Texas International Law Journal, 50*(2), 305–353

[57] Sullivan, C. (2013). In R. Jones & R. Moore (Eds.), *Information Technology and Traditional Legal Concepts.* Routledge. Retrieved from https://books.google.co.uk/books?id=YdCMAQAAQBAJ.

[58] European Commission. (n.d.). The European Cloud Initiative. Retrieved August 4, 2016, from https://ec.europa.eu/digital-single-market/en/%20european-cloud-initiative.

differing political and legal agendas across key players in the international arena and availability of other, more appropriate legal mechanisms for regulating the data industry. In light of this, it can be anticipated that a unilateral or multilateral CSR scheme, which picks up on the burgeoning ethics agenda would be the most viable option going forward.

A number of further recommendations for such a scheme can be made. To enhance the efficiency and impact of data financing across diverse jurisdictions, future mechanisms should ideally target social needs insufficiently met by current regulatory mechanisms. Future proposals should clearly establish the gap in social goods to be met, and the relative efficiency of the proposed mechanism. Governance regimes must also agree on common rules to calculation liability for contributions. Where schemes are trans-national, agreement can be expected to involve lengthy deliberations given conflicts between national legal regimes. Finally, attention must be paid to potential overlaps with existing industrial CSR schemes.[59]

Data financing can also be accomplished through expansion of existing CSR schemes, for instance through tax incentives for infrastructure investments by key actors in the digital economy. Well-designed mechanisms of the future will be responsive to the opportunities and shortcomings of existing CSR initiatives and related work carried out within data-intensive industries.

To enhance the efficiency and impact of data financing across diverse jurisdictions, future mechanisms should ideally target social needs insufficiently met by current regulatory mechanisms.

Internet.org: Connecting the World, But at What Cost?

Launched by Facebook in partnership with six telecommunications companies[7] in August 2013, Internet.org is a global initiative that aims to provide internet access to unconnected parts of the globe.[8] The core of Internet.org is Free Basics, a service that provides the poorest members of global society with access to a select range of websites without data charges. The included websites cover a broad span of content, including news, health, education, employment, and communication tools.

Although Internet.org claims to have brought more than 25 million people on line, the fact that the service covers only selected websites (including, for example, Facebook but not any of its main competitors) has attracted criticism from net neutrality advocates who worry that it distorts competition and undermines the fundamental openness of the Internet. This confrontation culminated in complete rejection of Free Basics by Indian regulators in February 2016.[9]

Thus, Internet.org represents some of the fundamental trade-offs inherent in this space: companies are ready to commit significant resources to projects that are potentially valuable for society, but will generally look for opportunities to earn a return, shelter themselves from competition, and minimise opportunities to free-ride on their investments. Whether this is a price worth paying is an important and open question.

[7] Samsung, Ericsson, MediaTek, Nokia, Opera, and Qualcomm.

[8] Internet.org. (2016). *Connecting the world.* Retrieved August 4, 2016, from https://info.internet.org.

[9] Ribeiro, J. (2016, February 8). Facebook's Free Basics prohibited in India. *Computerworld.* Retrieved from http://www.computerworld.com/article/3031152/internet/facebooks-freebasics-prohibited-in-india.html.

[59] It must be noted that CSR is a term used by the industry to describe altruistic and philanthropic initiatives. Some scholars disagree that it is an appropriate term, because many CSR schemes also advance business interests (e.g., the acquisition of more users). Facebook's Free Basic programme, for instance, contains an altruistic motive to spread Internet access, while also adding to the company's user base.

5 INDUSTRY ACCEPTABILITY

Key Questions

- What are the key concerns of data-intensive companies that relate to the four proposed data financing models?

- Which dimensions or features would make the models more acceptable to industry stakeholders, or less so?

Summary

- Industry attitudes are an essential element of the feasibility of innovative finance in the data economy. Beyond having sound social, economic, ethical, political, legal, and technical justifications, any mechanism proposed in the future must take account of the likely reaction of affected industry stakeholders.

- Based on unstructured feedback from industry stakeholders and the four project workshops, six types of concerns were identified in response to the four proposed data financing models. Together, they form an initial map of issues to be considered in the design of any future data financing mechanism for the data economy.

- The concerns address the effect on technological innovation, value for users, ownership of personal data, the size of stakeholder firms, bureaucracy and waste, and principled opposition to data financing from data-intensive firms.

- Given these concerns it is clear that for any scheme to succeed there must be an alignment of interests; compelling incentives must be created.

- A strong positive argument is needed for a multilateral implementation of a global internet subsidy, given the existence of comparable decentralised schemes.

- A privacy insurance for personal data processing raises a moral hazard issue: there is the risk that companies will be less incentivised to maintain data security if the insurance is sufficiently comprehensive.

- An attention levy might create incentives to craft alternative digital advertising models more resilient towards exploitation.

- Duties collected from revenue generated from open datasets can be of net benefit to the economy if they are used to support the increased availability of such datasets.

Analysis

Industry attitudes are an essential element of the feasibility of innovative finance in the data economy. Beyond having sound social, economic, ethical, political, legal, and technical justifications, data financing mechanisms must take account of the likely reaction of affected industry stakeholders. This chapter presents the key themes represented in unstructured feedback received from industry stakeholders in relation to the four data financing models explored in this report. It also highlights the incentives and disincentives that would be created by each of the four models.

5.1 Industry Concerns

Based on unstructured feedback from industry stakeholders and the four project workshops, six types of concerns were identified in response to the four proposed data financing models. Together, they form an initial map of issues to be considered in the design of future data financing mechanisms. The map provides guidance for structuring data financing partnerships with industry stakeholders.

5.1.1 Technological Innovation

Argument: Innovation drives the data economy of the twenty-first century. Innovation is an inherent good that must be supported. A levy on data may have a chilling effect on innovation. Depending upon the extent of the financing mechanism, existing technology companies may stagnate and start-ups may fail to form as rapidly and extensively as a result.

The collection and analysis of vast amounts of unstructured data have generated commercial opportunities that never before existed. Such opportunities accelerate the

creation of technology start-ups and new business models, as well as the improvement of existing products and services (through, e.g., value chain optimization or greater personalization). Ultimately, data-driven innovation can be leveraged to promote sustainability, resource efficiency, growth, and well-being.[60] The connection between big data and humanitarian action is, for instance, well argued: digital behavioural data has proven useful for detection of warning signs, real-time monitoring and feedback during crisis situations.[61]

All of this being said, innovation is not always beneficial. Not all innovation is positive. The term 'disruptive', often used hand in hand with innovation, has begun to acquire negative connotations (in part because it has been used so egregiously).[62] Although innovation can lower prices and increase the accessibility and affordability of products and services, it can also irrevocably damage other businesses. Online retailers are a good example: Amazon has put many companies out of business, from local bookshops to multinational chains like Borders. On a larger scale, innovation in the financial industry was a key factor of the 2008 economic crash, and new revenue streams generated in certain sectors (e.g., creative arts and travel) by digital innovation have not matched the scale of previous revenue streams.

Both the social goods and disruptive effects of data-driven innovation would be imperilled. With that said, while the effects of a data financing mechanism on innovation are undoubtedly important, the general value of innovation is a background concern. Planning of future mechanisms need not explicitly address the inherent value of data-driven innovation. Rather, if innovation is seen as desirable in the targeted industry sector, future mechanisms should be proposed in a manner that is responsive to varying capacity of firms to absorb new costs.

Proposals for a licensing scheme to extract personal data could, for example, imperil start-ups lacking sufficient revenue to afford the licence.

If innovation is seen as desirable in the targeted industry sector, future mechanisms should be proposed in a manner that is responsive to varying capacity of firms to absorb new costs.

5.1.2 Value for Users

Argument: Technology companies already create substantial value for society through the goods and services that they provide. The creation of this value already represents a fair contribution to society and fulfilment of any implicit social contract. Moreover, any chilling effects associated with data financing are likely to curtail the extent to which these services are offered in future, harming users.

Information technology companies generate significant value for their users: opportunities for social interaction, information discovery, and entertainment are three key benefits. Some of this value is economic, as users can themselves make money from participation in the digital ecosystem. Moreover, many companies help users to save money (by facilitating price comparison) or offer conveniences such as home delivery. But there are also significant intangible gains from the wide-array of services on offer—often at no (pecuniary) cost.

Thus, it is true that technology companies serve a valuable role in society and should not be subjected to per se denigration. But playing a valuable role in society should not absolve an organisation of the responsibility to contribute equitably to the social resources from which it benefits. Nor does it exclude an organisation from the responsibility to behave

[60] OECD. (2015). *Data-driven innovation: Big data for growth and well-being.* Paris: OECD Publishing. doi:10.1787/9789264229358-en.

[61] United Nations Global Pulse. (2012, May 29). "Big data for development: opportunities & challenges": A Global Pulse White Paper. Retrieved from http://www.unglobalpulse.org/BigDataforDevWhitePaper.

[62] The Economist. (2014, July 3). Negative externalities. Retrieved from http://www.economist.com/blogs/democracyinamerica/2014/07/disruptive-innovation.

in a non-exploitative manner. A carefully designed data financing mechanism would seek to address specific inequities or dubious practices along the lines discussed in Chapter 2.

The concern that a data financing mechanism might, nevertheless, harm society by robbing it of access to beneficial services currently provided by technology companies is valid. Thus, measures should be taken to minimize this harm. For example, a mechanism applied to activities that are socially harmful will, by reducing the prevalence of those activities, benefit rather than harm society. Lump sum levies are much less likely to introduce harmful distortions than are proportional unit taxes.

It is true that technology companies serve a valuable role in society and should not be subjected to per se denigration. But playing a valuable role in society should not absolve an organisation of the responsibility to contribute equitably to the social resources from which it benefits.

Lastly, it is important to consider the benefits, as well as the costs associated with a data financing mechanism. One reason why society tolerates distortionary taxation of any form is that the proceeds can be used to pay for valuable public services that benefit all—including tech industry stakeholders—to an extent sufficient to offset the social costs of the tax. Some distortion of trade in technology markets may be socially justified if the resources raised can be put to sufficiently good use.

5.1.3 Ownership of Personal Data

Argument: The use and sale of personal data is one of the biggest concerns for data-intensive companies because advertising and data transactions are key revenue sources. A transaction-based data financing mechanism may jeopardise the commercial viability of data processors if monetary costs or the effort required for compliance are high.

Data ownership remains nebulous in many legal frameworks. Often, terms of service stipulate that users retain ownership of the data they create, but also allow the service provider to use this information to improve their service and share with third parties. Companies can assemble digital profiles of users by purchasing data about their users from other companies —one's search behaviour on one platform can be related to one's exposure to advertisements on another. Such digital profiles can also be bought and sold amongst partner companies[63] and become powerful predictors of purchasing behaviour.

When accepting terms of service, users are seldom aware of these transactions. In recent years, however, more users have voiced negative feedback about their personal information being made available. Thus, there is justification for a social intervention here, and for proceeds raised to contribute to privacy protection technologies.

5.1.4 Scale of Stakeholders

Argument: Not all technology companies are created equal. Within the technology sector, there is a significant concentration of power and wealth. Thus, any innovative finance models that are proposed will have different effects on the most and least powerful players—regardless of the extent to which they process personal data.

It is important to bear in mind that the technology sector is highly heterogeneous, consisting of both large and small organisations with a wide variety of business models. A particular concern is that start-ups and SMEs will be most adversely impacted by the proposed schemes. They do not have financial resources to invest significantly in development, nor would they benefit substantially from the network effects of more Internet users. Small businesses depend on online advertisements to generate leads,

[63] Hildebrandt, M. & Gutwirth, S. (2008). *Profiling the European citizen.* Springer.

as they cannot afford large storefronts. The additional cost imposed by an attention levy might prevent them from acquiring enough customers to sustain their services. Moreover, because many start-up ventures use open data in their business model, a shared knowledge duty would stifle their formation in the first place. If such a duty were to be implemented, it should be designed to allow small companies to maintain free access to public data, while companies above a certain size would have to pay. These are important concerns, and the design of any data financing scheme should take into account the need to protect that openness and spirit of entrepreneurialism and experimentation that has been the source of much of the technology industries vibrancy.

One approach to dealing with this issue would be to target a data financing mechanism at a select group of larger organisations, or to operate a minimum organisation size threshold below which the mechanism does not apply. Besides protecting start-ups, focusing the mechanism only on larger organisations in this way may be justified for a number of reasons. Firstly, small technology start-ups are often more geographically localized than their larger peers and are less likely to engage in accounting practices that result in abnormally low tax contributions. Thus, such companies may be regarded as already making a fair contribution to the cost of public goods under the auspices of the normal business taxation regime. Secondly, several very large technology firms appear to enjoy a dominant position thanks to natural economies of scale. This implies a substantial flow of profits that may be viewed as unearned. A data financing mechanism on natural monopolies arguably has greater legitimacy than a tax on smaller businesses that compete purely on merits.

In other cases, the financing scheme would be designed to target particular harms such as negative externalities or exploitative behaviour. The attractive feature of such schemes is that they encourage firms (of any size) to reflect upon the effects that their choices have for others. These schemes would, by their nature, scale in line with the scope of the problematic behaviour. Take, for example, a corrective financing mechanism intended to disincentivise data maximalism. In order to deter a firm from storing more data than needed, it would necessarily be the case that storing more data implies a larger liability. Thus, a small firm that handles only a small amount of data would pay less than larger operation—commensurate with the fact that it exposes fewer users to less harm.

One approach to dealing with this issue would be to target a data financing mechanism at a select group of larger organisations, or to operate a minimum organisation size threshold below which the mechanism does not apply.

5.1.5 Bureaucracy and Waste

Argument: Creating an organisation to administer a new data financing mechanism would create bureaucracy and waste. Organisations often become as concerned with their own survival and growth as with their actual missions. Individuals within organisations also have their own agendas. Many existing organisations in the development sector are afflicted by these maladies. Sometimes the cure is worse than the problem it is meant to address.

The governance of any proposed compulsory or multilateral data financing model poses an issue in itself. How can industry stakeholders or the general public trust that the funds raised are being handled responsibly and efficiently? This is a perennial challenge in philanthropy and development aid. In economics, it is related to the problem of 'government failure': the notion that any government corrective to a market failure may itself be subject to a variety of failures. Besides actual failures, organisations responsible for others' funds must also avoid impressions of impropriety, creating another layer or complexity and potential waste.

A relevant example from the United States is E-Rate, a programme administered by the Universal Service Administrative Company (USAC). It is funded with levies imposed on

companies providing telecommunications services, and it distributes those funds to schools and libraries to help them obtain telecommunications services and Internet access. Over the years it has been subject to scrutiny over "waste, fraud, and abuse."[64] It is, however, subject to legislative oversight from the Federal Communications Commission, which has helped to bring the waste into light. It is likely that any data financing mechanism would likewise need credible and effective governance and oversight mechanisms to have a chance of receiving a favourable reception from the industry.

In data financing, there is also the novel possibility of using technology for some aspects of governance. Instead of rules expressed in legal code and implemented by bureaucrats, some aspects of a mechanism could potentially be governed by rules expressed in computer code and implemented by algorithms.[65] This could help increase transparency and reduce compliance costs, though it could also create complexity and introduce technological risks.

5.1.6 Principled Opposition

Argument: In the technology industry and especially in Silicon Valley, there is often a strong belief in self-regulating markets and an aversion towards government intervention. It is believed that social good is served best by free innovation and voluntary efforts. Any efforts to create data financing mechanisms based on compulsory law or taxation are therefore likely to provoke principled opposition from the industry even if the substantive challenges outlined above were addressed.

Libertarian and individualist ideologies have a long history in Silicon Valley, and are in many ways part of the technology industry's cultural DNA.[66] Innovation as a process involves spontaneity and breaking away from convention, and as such is ill amenable to central direction. However, in practice, many Silicon Valley companies accept that government has a role in enabling the conditions in which their innovation can flourish. For example, in 2014, Google, Facebook, Amazon, and other technology companies petitioned US regulators to mandate net neutrality to prevent cable companies from creating a tiered Internet.[67] It is also widely recognized that California's regulations against non-compete clauses in employment contracts contribute to Silicon Valley's innovativeness.

Innovation economists such as Mariana Mazzucato have also recently highlighted the historical and ongoing role of strategic government action in the success of Silicon Valley and other global innovation hotspots.[68] Fundamental innovations such as the Internet and GPS are rooted in strategic government investments and public-private partnerships. They serve as platforms for further private entrepreneurship. In some innovation hotspots, such as in the Nordics, the technology industry's relationship with government remains more openly collaborative than antagonistic.

One facet of individualism is that technology companies may not be as interested in voluntary but multilateral initiatives towards global good. Existing CSR initiatives are usually associated with a single company or entrepreneur. Besides any ideological preferences, this is simply a practical matter of maximizing the public relations benefits and other returns on investment. Data financing mechanisms based on individual voluntary action are thus likely to be the most palatable across the technology industry. But mechanisms where the government or

[64] U.S. Government Printing Office. (2005, November). Waste, fraud, and abuse concerns with the E-Rate program. Retrieved from https://www.gpo.gov/fdsys/pkg/CPRT-109HPRT24466/html/CPRT-109HPRT24466.htm.

[65] Lehdonvirta, V. & Robleh, A. (2016) Governance and Regulation. In: M. Walport (ed.), *Distributed Ledger Technology: Beyond Blockchain.* London: UK Government Office for Science, pp. 40-45. http://vili.lehdonvirta.com/wp-content/uploads/2016/05/Lehdonvirta-Ali-2016-Distributed-ledger-governance-regulation.pdf

[66] Barbrook, R. & Cameron, A. (1996) The Californian Ideology. *Science as Culture* 6(1): 44-72.

[67] The Guardian (2014) Google, Facebook and Amazon write to FCC demanding true net neutrality. *The Guardian* 8 May. https://www.theguardian.com/technology/2014/may/08/google-facebook-and-amazon-sign-letter-criticising-fcc-net-neutrality-plan

[68] Mazzucato, M. (2015) *The Entrepreneurial State.* New York: Public Affairs.

an international organisation takes the role of an enabler or even a strategic driver of an innovative platform are not entirely without precedent and understanding from industry quarters.

Innovation as a process involves spontaneity and breaking away from convention, and as such is ill amenable to central direction.

5.2 Incentives and Disincentives

Given all of the concerns raised in the previous section, it is clear that for any scheme to succeed there must be an alignment of interests; compelling incentives must be created. Such incentives can promote collaboration, as companies work with their competitors if the cause is mutually beneficial.[69] On the other hand, companies retain the capacity to operate autonomously, in their own self-interest. Incentives and disincentives created by the four innovative finance models are reviewed in this subsection.

A strong positive argument is needed for a multilateral implementation of a global Internet subsidy, given the existence of comparable decentralised schemes (e.g. Internet.org). Expanding Internet access is in the interest of most information technology companies. One incentive that would encourage more companies to participate in a global Internet subsidy is tax benefits. This raises a different issue, however: convincing national governments that money should be redirected from them to developing countries.

A privacy insurance for personal data processing raises a moral hazard issue: there is the risk that companies will be less incentivised to maintain data security if the insurance is sufficiently comprehensive. Some level of duty or care must be set as a prerequisite for receiving compensation in the event of a breach. This model could benefit smaller companies, as they are especially vulnerable to cybersecurity attacks. These companies would save a significant sum of money in the event of an attack, and the larger organisations that work with them would benefit as well.

The bigger incentive for all companies, of course, is maintaining a proactive reputation with regard to privacy. Edward Snowden's revelations in 2013 have triggered increased concern over not only government surveillance, but also the selling of personal information by data companies. According to a recent Pew Research Centre survey, ninety-one percent of Americans believe that they have lost control over the collection and use of their personal information; 'dataveillance' is an ever-expanding threat.[70] Companies like Facebook and Google have introduced highly granular privacy settings and easy-to-navigate privacy policies to increase transparency and control.[71] The funding of development of privacy enhancing technologies through an insurance scheme would provide a new channel for reputational gains.

Data financing mechanisms based on individual voluntary action are thus likely to be the most palatable across the technology industry.

The advertising-based model that has financed the Web over the past two decades is broken, and an attention levy might create incentives to craft alternative models more resilient to exploitation.[72] Similarly, a levy may contribute to the resolution of problems facing the current model, such as the use of bots to fraudulently gain ad revenue.

[69] A recent example of this is the Self-Driving Coalition for Safer Streets, a group formed by Ford, Google, Uber, Lyft, and Volvo to lobby for autonomous vehicles. See Hawkins, A. (2016, April 26). Google, Ford, and Uber just created a giant lobbying group for self-driving cars. *The Verge.* Retrieved from http://www.theverge.com/2016/4/26/11510076/self-driving-coalition-ford-google-uber-lyft-volvo-nhtsa.

[70] Rainie, L. (2016, January 20). The state of privacy in America: What we learned. *Pew Research Center.* Retrieved from http://www.pewresearch.org/fact-tank/2016/01/20/the-state-of-privacy-in-america/.

[71] See privacy.google.com (Google Privacy) and facebook.com/about/basics (Facebook Privacy Basics) for examples of comprehension-facilitating privacy policies.

[72] Zuckerman, E. (2014, August 14). The Internet's original sin. *The Atlantic.* Retrieved from http://www.theatlantic.com/technology/archive/2014/08/advertising-is-the-Internets-original-sin/376041/.

Advertisers are being exploited because of the fake pageviews generated by such bots; they are already paying more than they should. The cost per thousand fake impressions adds up: almost a quarter of online video ads are viewed by bots, which contributed to a loss of over $6 billion for the advertising industry in 2015.[73] A levy would provide ad publishers and networks with an incentive to address the issue.

Finally, duties collected from revenue generated from open datasets can be of net benefit to the economy if they are used to support the increased availability of such datasets. Those that profit from open data can receive more open data as a result of the duty. Alternatively, proceeds could support increased privacy and security protections needed to make such datasets suitable for public use. The economic gains of such increased availability can be demonstrated to outweigh the costs of the duty.

It is clear that for any scheme to succeed there must be an alignment of interests; compelling incentives must be created.

5.3 Conclusion

Through unstructured feedback and discussion with industry stakeholders, a number of potential concerns and incentives for data financing mechanisms were identified. For any mechanism to be implemented, many concerns involving issues such as innovation, free markets, company size, traction, and bureaucracy must be addressed. Incentives for reputation, transparency, and open data can be leveraged.

Introducing a data financing mechanism to create different incentives in the marketplace might strike many as a form of micromanagement. The mechanism might be more acceptable if it is logical (e.g., money raised by a privacy insurance being used to fund privacy-enhancing tools). Credible and effective governance and oversight mechanisms are also key for mechanisms not controlled by industry partners, including potential automated governance by algorithms or within existing platforms for data transactions (e.g. a per transaction tax). Rather, the extent to which the mechanism can create incentives for contributing stakeholders is compelling.

Data financing in general must strike an appropriate balance between preserving innovation and achieving social goods through distortion of trade in data-intensive markets. Costs should ideally be realistic insofar as they do not render a firm's business model unviable, particularly for start-ups. Lump sum levies are much less likely to introduce such distortions than proportional unit taxes. Mechanisms can also target specific large data-intensive firms or operate with thresholds based on size or revenue.

[73] Rushe, D. (2014, December 9). Nearly 25% of "people" viewing online video ads are robots used by fraudsters. *The Guardian.* Retrieved from https://www.theguardian.com/technology/2014/dec/09/online-ads-robot-fraudsters.

6 CONCLUSION

In the preceding chapters, we assessed the feasibility of data financing from multiple perspectives. We found that there are sound economic, social and ethical arguments that could be used in justifying a variety of data financing models. In other words, data financing would not – and should not – simply be about redistribution for the sake of redistribution, but about correcting economic and moral problems arising in the data economy. We illustrated these justifications with four data financing models: a global Internet subsidy, a privacy insurance, a shared knowledge duty, and an attention levy. We also saw that good arguments can be mustered against many types of data financing, so any such initiative would have to pay close attention to ensuring that its existence can be reasonably justified.

As soon as one considers how any data financing mechanism might be implemented in practical and technical terms, a very complex landscape of challenges and potential solutions is revealed. Much of this complexity is due to the nebulous nature of 'data' as a concept and commodity. A very basic requirement is that administering a data financing mechanism, especially a compulsory one, requires a reliable and verifiable method of calculating dues. The most obvious approach is to use the volume or quantity of data as a measure on which dues are based. However, both are extremely context-dependent measures, with the consequence that the results could be highly arbitrary when applied across a variety of firms and industries. Volume and quantity might be more feasible measures for models that target very specific industries or verticals.

Another approach to calculating dues in a data financing mechanism is to measure the value of the data rather than the data itself. For certain limited types of data that are bought and sold, it is possible to observe a market value. However, much of the data trade happens 'over the counter' (OTC), with the consequence that transactions are not observable; this is the case with secondary uses of personal data collected from users, for instance. In the financial services industry, regulators have in some cases forced all OTC traders of a product to move to a public marketplace in order to make the market observable and thus amenable to regulation and taxation. The applicability of this idea to data markets was not investigated in this study, but is potentially worth examining in future studies. It may also be possible to use information disclosed in company financial reports to produce rough proxies of the value of firms' data-intensive activities.

The question of practical and technical challenges is interlinked with the question of how a data financing scheme could and should be organized in political and legal terms. For instance, if a data financing mechanism is voluntary rather than mandatory, it may be possible to avoid many of the practical and technical challenges, as long as contribution levels can be set in agreement between members. A unilateral or multilateral CSR scheme may be the most politically and legally feasible way of structuring a data financing mechanism. Indeed, many leading information technology companies already contribute to social good through CSR initiatives, though some of these initiatives are also aimed at further entrenching the company's own market position in the global data economy. There might be opportunities to build on these individual initiatives to create multilateral industry-spanning efforts that would deliver social good whilst moderating individual companies' attempts at market dominance.

There is also some international political momentum that could be used towards establishing a compulsory data financing model, focused on issues of Internet development and data protection. Growing concern over the ethics of big data, data analytics and machine learning may create a window of opportunity for establishing mechanisms related to these topics. There are, however, a number of significant political and legal barriers that would have to be overcome before any compulsory data financing model could be introduced. The key obstacles are the lack of a legitimate governance mechanism, differing political

and legal agendas across key players in the international arena, as well as the availability of other, possibly more appropriate legal mechanisms for regulating the data industry.

Sound governance and oversight are also among the issues that private sector stakeholders are likely to pay attention to in assessing any proposal for a data financing mechanism. This applies to compulsory as well as voluntary mechanisms. Private sector stakeholders and especially technology firms are also likely to call attention to the effects that data financing would have on innovation. Data financing might be particularly problematic for start-up companies experimenting with technologies and business models. Lump sum levies are much less likely to be disruptive than proportional unit contributions. Besides the contribution as such, compliance costs in the form of additional reporting and bureaucracy should be absolutely minimized. Mechanisms can also target specific large data-intensive firms or operate with thresholds based on size or revenue.

6.1 Scale of impact

The final question that we have not addressed is financial impact. Assuming that a data financing mechanism can be successfully implemented, what scale of impact could it feasibly have? A number of studies have estimated the direct value added of Internet-related activities in various developed economies to be between 0.8 and 13.8 percent of the gross domestic product (GDP).[74] Most estimates place the value added in the vicinity of 3-4 percent of GDP. The studies use varying definitions, but Internet-related activities are typically understood to include activities on the Internet (e.g. online publishing, online advertising, ecommerce) as well as activities supporting the Internet (e.g. Internet service provision, hardware manufacturing). Value added is defined as the value of the outputs minus the value of the inputs that went into their production; on the firm level it is parallel to profit.

Beyond the direct value added of Internet-related activities, the Internet and the exploitation of data are also expected to be having indirect effects on productivity across industries. Quantifying these indirect or dynamic effects is notoriously difficult; one OECD study concludes, with various qualifications, that up to 7.2 percent of US gross domestic product in 2011 was generated thanks to the Internet. Other studies report figures of similar magnitude. At the same time, the United Nations' development aid spending goal is 0.7 percent of a donor country's gross national income; most donors are spending substantially less. In other words, if it were possible to direct a fraction of the value generated in the data economy into global development, the impact would be close to the same order of magnitude as all existing development aid. And unlike national development aid budgets, the value generated by the data economy is poised to grow.

However, the total value of the global data economy only gives us an absolute upper bound. Actual yields would depend on what kinds of financing mechanisms could be justified and implemented. To give some idea of the potential scale of different financing mechanisms in the global data economy, we produced estimates of the yields of the four models discussed in this report. The estimates are 'Fermi estimates' designed to get at the approximate scale of the phenomenon with few inputs. They are not based on an exhaustive search of data sources, and emphasise transparency over sophistication. This level of detail is commensurate with the level of maturity of the models, more detailed estimates requiring more detailed models. It is important to note that the methods used to produce and present the estimates are not intended to suggest how any actual mechanisms should work. For instance, when an estimate is based on figures on a particular category, it is not to suggest that the incidence should necessarily fall on that category; and when terms such as 'tax base' are used to explain results, it is

[74] OECD (2013) Measuring the Internet Economy: A Contribution to the Research Agenda. *OECD Digital Economy Papers* 226. Paris: OECD Publishing. http://www.oecd-ilibrary.org/science-and-technology/measuring-the-internet-economy_5k43gjg6r8jf-en; Dean, D. et al. (2012) *The Internet Economy in the G-20.* London: Boston Consulting Group. https://www.bcg.com/documents/file100409.pdf

not to suggest that the legal implementation should necessarily be a compulsory tax.

6.1.1 Attention levy

The largest possible tax base of an attention levy against online advertising is equal to the industry's global revenues, estimated at around $170bn per year.[75] To estimate the yield that could be generated from this tax base, we can use the rate that James Tobin used as an example in his proposal for a currency transaction tax: 0.5%. This is largely an arbitrary number, but exemplifies an order of magnitude that on the one hand can conceivably be levied without bankrupting business, and on the other hand is not economically insignificant. It may seem small compared to, for instance, typical corporate tax rates, but note that the rate is being applied here to gross revenues rather than profit. It is certainly not the only conceivable rate, nor the only conceivable way of assessing such a levy, but a more detailed examination is contingent on the particulars of a given scheme and thus goes beyond the scope of this report.

With the Tobin rate, the upper-bound estimate on yield from an industry-encompassing attention levy is $850m. A more focused levy might be aimed at the, arguably more intrusive, market for display advertisements, with a market size of around $63bn[76] — implying a maximum yield of $315m from a 0.5% levy. As a third alternative, a levy might be targeted at a small number of major ad platforms that serve as bottlenecks for the online advertising industry. The relevant dollar order of magnitude for ad revenues on a major platform is between hundreds of millions to (low) tens of billions. Thus, a levy of 0.5% would generate an annual yield in the millions or tens of millions of dollars per-included platform.

Note that one motivation for imposing an attention levy is that, by making ads more expensive, it reduces the volume of spurious advertising and therefore mitigates excessive trespass on users' attention. But this change in price and quantity implies a change in revenues that will affect the above calculations. Fortunately, there is reason to believe that these dynamic effects will be relatively small. For example, Choi (2014)[77] provides estimates from the Korean search advertising industry that imply a price-elasticity of around -0.23. This means that a 0.5% increase in ad prices (caused by a 0.5% attention levy) would lead to industry pre-levy revenues increasing by a factor of (1+0.005)(1-0.0012)=1.009—or 0.9%. Thus, making this correction implies that a 0.5% levy on total global online ad revenues would result in a levy yield of $857m, rather than $850m.

6.1.2 Privacy insurance

In its 2015–2016 reporting year, the UK's Information Commissioner's Office levied fines of £2.529m ($3.321m) for breaches of the data protection act.[78] This amounts to £0.04 ($0.05) per-capita per-year in fines. Extrapolating, for example, to the European Union as a whole, a similar per-capita incidence of fines would yield an annual income of £29m ($39m). The imminent introduction of the GDPR within the EU is set to significantly increase the level of fines imposed for data breaches (to 4% of annual revenue), suggesting that the appropriate level of penalty may be much higher still.

Another approach to the issue is to put aside issues of liability/negligence and to view the problem more literally as one of insurance. IBM/Ponemon estimate that, in the US, each record stolen in a data breach implies an average cost of $217 for the organization concerned.[79] The same source estimates that the average US data breach involves 28,070

[75] Brouwer, B. (2014) Global Mobile Ad Spending Will Hit $159 Billion By 2018. *IMDB* 24 December. http://www.imdb.com/news/ni58132714/

[76] Lunden, I. (2014) Internet Ad Spend To Reach $121B In 2014, 23% Of $537B Total Ad Spend, Ad Tech Boosts Display. *Techcrunch* April 7. https://techcrunch.com/2014/04/07/internet-ad-spend-to-reach-121b-in-2014-23-of-537b-total-ad-spend-ad-tech-gives-display-a-boost-over-search/

[77] Choi, D. O. (2014) *Internet Advertising with Information Congestion.* Mimeo. http://www.webmeets.com/files/papers/res/2014-phd/259/congestion.pdf

[78] ICO (2015) *Information Commissioner's Annual Report and Financial Statements 2015/16.* Cheshire: Information Commissioner's Office. https://ico.org.uk/media/about-the-ico/documents/1624517/annual-report-2015-16.pdf

compromised records, while the Identity Theft Resource Center tracked 781 breaches in 2015.[80] Together, these figures imply an annual cost to businesses of around $4.8bn. Thus, an actuarially fair universal public insurance scheme that compensated firms against the costs of data breaches would have a value of around $4.8bn a year. If the scheme deviated from the actuarially fair price to include a 0.5% surcharge, then it would yield a surplus of around $24m/year that could be committed to other projects. As is usual in insurance markets, offering full insurance creates the potential for moral hazard. This could be mitigated either by offering partial insurance (and scaling the size of the insurance scheme accordingly) or by making pay-outs contingent on non-negligent conduct.

6.1.3 Global Internet subsidy

It has been estimated that, by the end of the decade, the economic exploitation of personal data will contribute a benefit of €330bn/year for European organizations.[81] If 0.1% of this value could be captured, the result would be a €330m annual fund for global infrastructure (or for development activities more generally). As an example of the impact that this level of funding can have, the Private Infrastructure Development Group used $375m (€340m) of funding to launch the SEACOM undersea cable project, which has been credited with reducing Africa retail bandwidth costs by 67% and enabling a significant increase in Internet penetration in the region.[82] If we go beyond personal data and consider the data economy as a whole, the potential tax base is much larger, as discussed in the following section.

6.1.4 Shared knowledge duty

The potential justifications for a shared knowledge duty relate to the notion of Internet companies deriving revenue from data shared by users, publishers, governments, and others. The potential scale of such a duty must therefore bear some relation to the revenues of these companies. The total annual revenues of the world's 20 top-earning Internet companies were approximately $305bn in 2015. This is based on Wikipedia's definition of an Internet company as a company that does the majority of its business on the Internet, excluding Internet service providers and other information technology companies.[83] It includes social media companies whose businesses are based on user contributions, but it also includes ecommerce firms who deal in physical goods with help from user data. If we assume that half of the revenues of this group pertain to the exploitation of shared knowledge, the potential tax base is $152.5bn. Applying the Tobin rate of 0.5%, we are left with an estimated yield of $762.5m. This does not account for any dynamic effects, such as reduced activity due to increased costs or increased activity due to proceeds being invested into growing the Internet's user base.

Extending the hypothetical tax base beyond the top 20 Internet companies increases the estimated yield. Gartner estimates that worldwide IT spending in 2015 totalled approximately $3.4tn,[84] or about 11 times the revenues of the top 20 Internet companies. However, this figure includes hardware manufacturing, Internet service provision, telecommunications, data centre provision, and other goods and services whose connection to the exploitation of shared knowledge is at best indirect. Assuming

[79] Ponemon Institute (2015) *2015 Cost of Data Breach Study: Global Analysis.* Ponemon Institute Research Report. https://nhlearningsolutions.com/Portals/0/Documents/2015-Cost-of-Data-Breach-Study.PDF

[80] ITRC (2016) Identity Theft Resource Center Breach Report Hits Near Record High in 2015. *Identity Theft Resource Center* January 25. http://www.idtheftcenter.org/ITRC-Surveys-Studies/2015databreaches.html

[81] BCG (2015) *The Value of Our Digital Identity.* London: Boston Consulting Group. http://www.libertyglobal.com/PDF/public-policy/The-Value-of-Our-Digital-Identity.pdf

[82] PIDG (2016) *Seacom Undersea Cable.* Surrey: Private Infrastructure Development Group. http://www.pidg.org/impact/case-studies/seacom-undersea-cable

[83] Wikipedia authors (2016) List of largest Internet Companies. Wikipedia, retrieved 10 August. https://en.wikipedia.org/wiki/List_of_largest_Internet_companies

[84] Gartner Market Databook, 2Q16 Update. https://www.gartner.com/doc/3361617

that as much as a quarter of the $3.4tn is attributable to the exploitation of shared knowledge results in a tax base of $850bn. Applying the Tobin rate of 0.05% to this base results in an estimated yield of $4.25bn per year, ignoring dynamic effects. Extending the hypothetical tax base beyond the top Internet companies thus increases the yield, but not by orders of magnitude. The revenues of the top four firms alone – Amazon, Google, Facebook, and Tencent – amount to $216bn or about 25% of our extended tax base.

6.2 Concluding remarks

Data financing is a provocative idea: it asks us to consider how private gains from the growing data economy could be directed to global social good, which faces substantial funding gaps. This report suggests that there are many obstacles on the way of implementing a data financing mechanism in practice, the least of which is not the difficulty of pinning down the concept of 'data'. At the same time, the tentative estimates presented above suggest that the impact of a successful mechanism on global good could be very significant. On this basis, we cautiously recommend further attention to the idea, with the understanding that rapid achievements are unlikely.

A suggested next step to move data financing from an idea towards implementation is conducting further studies that narrow down the scope by taking a sectoral approach, focusing on a specific industry, vertical, or form of data or data processing. Technical solutions or other practical means (e.g. financial reports) to establish a data financing mechanism in particular sectors can also be examined. Alternatively, further studies could focus on particular legal structures, such as a multilateral consortium of Internet companies, or an international legal instrument building on wider political momentum. Consideration can be given to the activities that can be brought under the umbrella of particular legal structures. Narrowing down the scope would allow studies to examine in detail the applicable subset of the many issues outlined in this report, perhaps developing concrete solutions. Studies could perhaps also approach the topic from an entirely different direction, asking what global problems there are that could best be addressed by data financing, considering the type and scale of problem, and any linkages to the global digital economy.

This report provides a first step in the assessment of data financing as a mechanism for social good in the data economy. To use the language of the technology industry, data financing is a 'moonshot'—a radical idea whose success is far from certain, yet addressing such a huge problem that it is tempting to at least give it a try.

Richard Wilcox
Special Advisor
Humanitarian Finance
United Nations World Food Programme

AFTERWORD

Big data is the basic resource of the new economy. Data are to this latest, fourth, industrial revolution what coal was to the first. Data as a global public asset can also help eliminate the scourge of extreme poverty at no additional cost to consumers and taxpayers.

Big data is creating massive value to the world economy. It powers product design, market strategies, targeted advertising and a host of other things that, taken together, drive the modern economy.

So valuable is this data that companies collect and store vast amounts of it for purposes they themselves sometimes do not yet understand. They extract it, and they store it, and the fact that they have done so is believed by investors to add to their value, on the assumption that a lucrative use for it will be found.

More than two thirds of this data is personal data—that is, it comes from you and me. It is the now necessary and unavoidable electronic traces of our everyday activities. It is gathered by data-maximizing companies—companies that generate value primarily based on the collection and use of personal data. Some of these data-maximizing companies are obscure, but many are household names, including Google, Facebook and Uber.

Data are inevitably created about you. It is impossible to function as part of the data-driven economy without your data being collected and used by data-maximizing companies—email, navigation, search, and so on—have become irreversible necessities of our lives.

Data—our data—has created the most valuable enterprises in human history. If we harness this resource justly to fund development, it could transform the lives of the poorest.

It is a choice. Either, collectively, we choose to allow our data to be harvested with little say in how it is used to benefit global society, or we put in place a system, as we do for every other resource, that will benefit the common good, and provide some resources for those left behind in extreme poverty.

It will not be simple. Citizens find themselves in a position of extreme asymmetry relative to the data extracting companies, making a fair market exchange between data maximizing companies and individuals implausible. Moreover, the power of individuals—those from whom the data are extracted—is dispersed, while the power of those who gather the data is concentrated. As an individual you are not well positioned to assess the value of the data you generate or to bargain with the companies collecting it.

The difficulties do not stop there. Data travels instantaneously across international borders, confounding the logic of states and tax jurisdictions that shaped the old economy.

What is to be done?

As the Oxford Internet Institute's study shows, a more just distribution is possible if numerous barriers can be overcome. Given data maximizing companies' market valuations, the cost of the extracting services

compared to the value of the data extracted, and the ever decreasing cost of storing and exploiting data, well calibrated international data financing mechanisms could be constructed.

This is not an idle moral argument. Commissioned by the Rockefeller Foundation, the Oxford exercise was an attempt to understand how the regular and accepted logic of fiscal policy can be applied to data extraction and processing. It is an early effort to assess how the creators of the world's most profitable resource can have a greater say in how it is used to benefit global society.

As proposed by Rufus Pollock of Open Knowledge International, a key analogy can be drawn between the environmental movement's rise in response to the first industrial revolutions and our present efforts in response to this fourth industrial revolution. In our understanding of the data economy's impact we are at a point today similar to the public's understanding of the previous revolutions' impact prior to Rachel Carson's 1962 *Silent Spring* giving the environmental movement its footing.

In today's data economy there are growing concerns about privacy and competiveness akin to the 1960s concerns about pesticides and pollution, but no realization yet of the new data economy's transformative social and political impact and opportunities.

We cannot yet claim to fully see the new data economy's social and political impacts, but I hope this study can provide an evidence-base for a global social movement to try and guide the data economy in socially beneficial directions—a movement which for the carbon economies came very late, perhaps too late, for our planet's climate. The European Union could lead the way by including a data extraction levy in its plans for the Digital Single Market by 2020. The United States, the EU and other developed countries could show openness to negotiating a well-governed international Internet and data regulatory and revenue system that would promote rather than restrict the free flow of data. The next UN Secretary-General will need to provide leadership on par with his or her predecessor's on climate change if we are to harness this new global resource to eliminate extreme poverty.

NOTES

NOTES