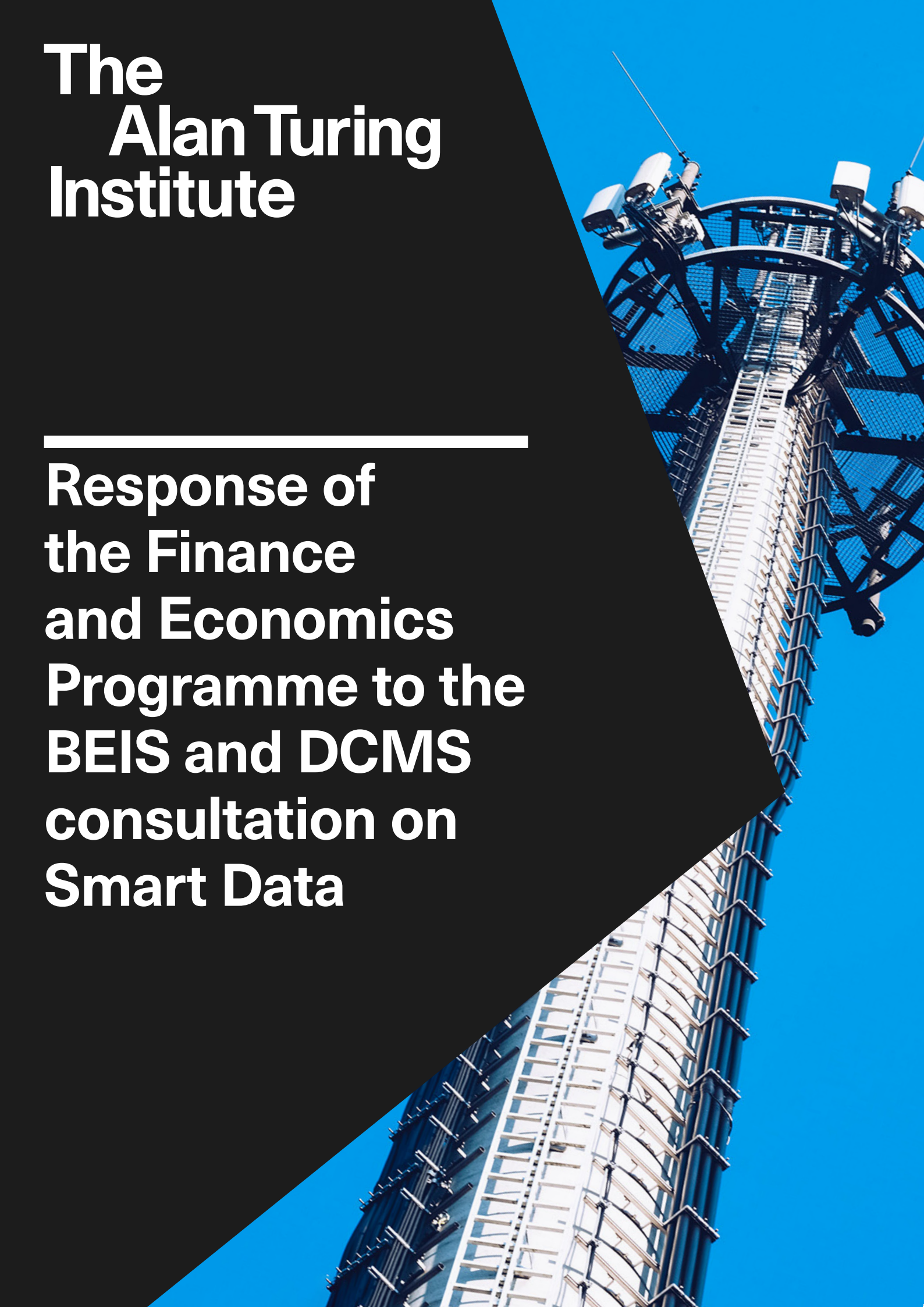
The Alan Turing Institute

Response of the Finance and Economics Programme to the BEIS and DCMS consultation on Smart Data



Response of The Alan Turing Institute's Finance and Economics Programme to the Department for Business, Energy and Industrial Strategy and Department for Digital, Culture, Media and Sport consultation on *Smart Data: Putting Consumers in Control of their Data and Enabling Innovation*

Introduction

This document provides the response of the Finance and Economics Programme at The Alan Turing Institute, to the Department for Business, Energy and Industrial Strategy and Department for Digital, Culture, Media and Sport consultation titled *Smart Data: Putting Consumers in Control of their Data and Enabling Innovation*.

The response combines the perspectives of several researchers from across the Institute and those in its wider university network. We are grateful for the contributions from the following researchers:

- Bonnie Buchanan (Head of Department of Finance and Accounting, University of Surrey)
- Jonathan Cave (Turing Fellow)
- Jon Crowcroft (Turing Fellow)
- Peter Grindrod (Professor of Mathematics, University of Oxford)
- Richard Gomer (Research Fellow, University of Southampton)
- Irene Ng (Turing Fellow)
- Florian Ostmann (Policy Fellow)
- Elena Simperl (Turing University Lead, University of Southampton)
- Anya Skatova (Turing Fellow)
- Melanie Smallman (Turing Fellow)
- Philip Treleaven (Strategic Advisor, Finance and Economics Programme)
- Johanna Walker (Web Science Institute, University of Southampton)

This response was prepared by Anastasia Shteyn, Programme Manager, Finance and Economics Programme.

Some discussion points in this document were informed by the National Data Strategy Roundtable hosted at The Alan Turing Institute on July 22, 2019. This roundtable was organised by The Alan Turing Institute's Public Policy Programme and the National Data Strategy Team at DCMS. This document, however, forms a separate response to the outcomes of the Roundtable.

The Alan Turing Institute (www.turing.ac.uk) is the UK's National Institute for Data Science and Artificial Intelligence (AI), headquartered at the British Library in the heart of London's vibrant Knowledge Quarter. Our aim is to be a world leader in data science and AI research and innovation.

The Finance and Economics Programme brings together leading experts in data science, machine learning, finance and the social sciences, from both academia and industry to tackle the most challenging questions by producing world-leading research with significant impact. The Programme works closely with the industry to exploit the potential of new technologies in the financial sector and economic research, and to position the UK as the leader in these areas.

This response will be split into two sections: the first will make comments on the overall vision outlined on the consultation; the second will address the specific questions listed in the consultation document.

Section 1: Overall comments and observations

The researchers see high potential in developing a consistent and coordinated approach to data sharing and consumer friendly innovation. The three themes below summarise the key high-level observations and further considerations beyond the specific consultation questions outlined in the vision document.

1. Clarity of definitions and scope of Smart Data

The whitepaper describes Smart Data as "easily and instantly accessible to consumers and be able to be safely and securely transferred to third party services who can use this data to provide innovative services for consumers." The contributing researchers would welcome greater detail around the definition of Smart Data, that goes beyond properties and affordances. Specifically:

1.1. The definition should focus on the provision of data to data subjects (users / consumers), as well as the provision of data by the data holder to Third-Party Providers (TPP). The legal and economic ramifications of this are outlined in Section 2, Question 1, based on the example of Open Communications.

1.2. The vision document states that Application Programme Interfaces (APIs) can be used to "share data securely, but only once the consumer has verified their identity." The researchers would welcome further examination of the assumptions that underpin this statement.

- TPPs can run analysis on data, but not be given the data itself, in turn decreasing risks associated with data privacy. This can be achieved using a number Privacy Enhancing Technologies (PETs) available today, such as homomorphic encryption.
- It may be enough for the user to prove they have the right credentials to access, download, or alter / remove the data. It's important to distinguish between a general-purpose identity infrastructure and a technical capability such as this one.
- Good data governance goes well beyond the notion of data control at the point of sharing; it involves ongoing monitoring and consent, understanding legal and economic rights of parties to contract for its use, reuse and sharing, as well as deletion of data after it's moved. PETs can be applied to all stages of storage, movement and processing. In order to define the precise scope of Smart Data initiatives, it's crucial to start with a baseline set of objectives of "good" data sharing and governance.
- There are technical and regulatory ways of realising the goal of organisations to "share data securely." For example, GDPR and effective fines can be used to enforce good behaviour, but this will likely entail requiring organisations to spend resources monitoring for exfiltration. At the other extreme, one can imagine requiring Data Rights Management (DRM) / container technology enforcing copyright protection on behalf of data subjects.

1.3. We would welcome a tighter definition for a Third-Party Provider (TPP) that considers what:

- Engenders user trust in a TPP
- Liabilities a TPP has
- Recompense is available if trust is breached

Furthermore, Jonathan Cave advises that there is inherent complexity associated with chains of data transfers and imperfectly understood networks of relationships. A data subject's interaction with a business generates data, and this data may be passed to other organisations directly (with consent) or indirectly – for example in an anonymised form or processed into a model. While the chain itself may be invisible, it has a continuing effect on what offers are provided or denied to the data subject. It is therefore important to think of how businesses can engender trust in their business models with the consumers. We further discuss the potential limitations of the consent approach in Part 2, Question 12 of this response.

2. National Data Infrastructure

Connected to the scope of Smart Data initiatives, we would encourage the Government to think about the role it would like to play in setting up the supporting data infrastructure.

2.1. Philip Treleaven stresses the importance of investing in such an infrastructure. One important reason is to mitigate the risk of losing the UK's large services sector to another economy that is the first to build digital marketplaces for financial and professional services. He believes that for any industrial company, the key to future innovation and wealth creation is marshalling its data resources and building the next generation data infrastructure.

2.2. There are currently several industry projects focused on developing a technological solution. They commonly rely on secure blockchain environments and executable smart contracts to allow organisations to access and process data in a permissioned way. Some of the models to be considered include a 'data trust' model, for example the ODI Data Trusts, the federated model that relies on person-owned and controlled "HAT microservers", for example the Hub of All Things project, and the blockchain models. All have unique advantages and challenges that need examination.

3. The balance between innovation and regulation

The title of the consultation obscures potential tensions between the goals of consumer control over data and enabling innovation: giving consumers control over their data is likely to preclude certain uses of consumer data that would be conducive to innovation. We would welcome clarification on the relative weights that the proposed smart data strategy assigns to these two goals and on how conflicts between these goals are meant to be addressed.

3.1. Philip Treleaven believes the data regulator needs to be proactive and encourage innovation. He points out that the Fintech community in London was imperative to the success of Open Banking; other sectors may require nurturing of the start-up ecosystems. Some examples include offering regulatory sandboxes and organising Tech Sprints to look at specific issues such as privacy and trust, for example the good practice set by the FCA for Open Banking.

3.2. Melanie Smallman advises that research into public attitudes suggests that public are 'resigned' to loss of privacy but are not happy about it. There is a risk in not getting consent and regulation right: innovation may be stifled by people 'opting out' of all forms of data collection and sharing.

3.3. Several contributing researchers highlight that data ethics underpin every level of the conversation around data governance and regulation. There are many projects and

initiatives at the Turing pertaining to data ethics and the associated practical guidance.

The Institute's Public Policy Programme has a dedicated team of researchers working alongside policy-makers to develop the ethical foundations for using data science and artificial intelligence in the public sector. In 2019, the Programme partnered with the Office for Artificial Intelligence and the Government Digital Service to produce guidance on the responsible design and implementation of AI systems in the public sector. The guide, Understanding artificial intelligence ethics and safety, is the world's most comprehensive guidance on AI ethics and safety for the public sector, offering concrete measures to counteract potential harms[1].

The Programme also collaborates with regulators in the UK on AI explainability and transparency. The Programme is working with the Information Commissioner's Office (ICO) to develop guidance to assist organisations in explaining AI decisions to the individuals affected.[2] The Programme also recently announced a new collaboration with the Financial Conduct Authority (FCA), on a research project that will examine current and future uses of AI across the financial services sector, analyse ethical and regulatory questions that arise in this context, and advise on potential strategies for addressing them. Finally, together with the Finance and Economics Programme, the Public Policy Programme recently hosted a conference on AI Ethics in Financial Sector, with keynote speeches from the FCA and the ICO, attesting to the Turing ongoing commitment to inform policymaking on ethical issues in the use of data technologies.

The Turing also has an active Data Ethics Group. Made up of academics specialising in ethics, social science, law, policy-making, and big data and algorithms, the Data Ethics Group drives the Institute's research agenda in data ethics and works across the organisation to provide guidance on ethical best practice in data science. The Group works in collaboration with the broader data science community, supports public dialogue on relevant topics, and sets open calls for participation in workshops, as well as public events.

[1] https://www.turing.ac.uk/sites/default/files/2019-06/understanding_artificial_intelligence_ethics_and_safety.pdf

[2] Interim report: https://ico.org.uk/media/about-the-ico/documents/2615039/project-explain-20190603.pdf

Section 2: Response to Consultation Questions 1-15

Enabling data driven innovation in consumer markets

1. Do you agree with the proposed objectives and expected benefits of Open Communications? Are there any other benefits or risks that we should consider?

1.1. Irene Ng believes that open communications has always been the ethos of the Internet. Companies can choose to share data with one another, with consent, through API calls, without a need for incentives to do more. She would welcome greater clarity on how the new initiative would be different from the status quo.

1.2. The vision document states that it wishes to be "Introducing an Open Communications initiative that will require communications businesses to provide consumers' data to Third Party Providers at the consumer's request." Irene considers this statement from the perspective of law and economic incentives and standards:

"First, what is the *legal* difference between the "consumer's request to give their data to third party providers" and "a third-party provider asking a consumer for consent to transfer data from a source", the latter currently allowable and unregulated. For example, the intention of the first could be data subject access request for the source to give personal data to a personal data store / account provider so that individuals can control the use of their data; the intention of the second could be an app asking for your profile photo on Facebook to show one how one might age.

Legally, economically and technologically, is there a difference between the two examples?

Legally, if there is a difference, then I believe the current law does not provide for it, nor will it separate a well-intentioned personal data store provider from an app set out to "hoover-up" personal data.

Economically, the value of data is not at rest but when it's shared and used. I believe data rights could be clearer and could be strengthened for data mobility, which incentivises opaque practices and therefore high transaction costs. Consent fatigue can lead to malpractice. Data that moves away from the source structure can be restructured, de-identified and re-identified to circumvent GDPR, often in real time and on demand, making it an economic asset that is freely disseminated without any regard for the consequences. Practices such as these may also be outside the jurisdiction of the UK, which is already the case for the ad tech market, with firms that specialise in de- and re-identification outside of GDPR jurisdictions. Is the Smart Data

function aimed to boost this sector? Without clearer data subject rights, the notion of "control" may become inconsequential, and apps can purport to give control without real data rights. If there is no motivation to create more or better data rights and only give "control", the eight rights under GDPR already provide for this to some extent, and companies that provide personal data store/control services show that the market is functioning although success at scale is not evident. The economic question is therefore one of "will giving consumers more control make the market function better" or is the issue one of a different set of data rights for consumer. We argue it's the latter.

Creation of a separate "TPP" construct, which is different from a general "application," suggests that the market is non-functioning in terms of personal data handling. There is a potential danger in that the very economic asset for open comms ceases to become one, the moment it is regulated.

The non-rivalrous and expansionable nature of data means that it can sit and be used concurrently in multiples places. Since data is not a fungible asset, the rights to use, share, control and process that data would depend on where it sits and who has rights to the "vessel" within which it resides. If personal data is at rest in centralised systems, it would be subject to GDPR. If it rests in de-centralised or federated systems, the way it's legally treated would depend on the type of vessel (e.g. mobile phone, PC) and the control of that vessel.

There is also a danger of personal data becoming non-personal data. Firms are already creating data structures where the data of activities is separated from user identity. There is a risk that a Smart Data function results of Smart Data may consist of solely people's names and email addresses and nothing else if apps and websites game the regulatory function.

I think the current objectives are too broad. Is it good when personal data is "fresh", "ethically sourced", used quickly, expires quickly and is then swiftly deleted? If so, how can this be enforced? And if enforced, how to incentivise apps not to leave the UK jurisdiction to do otherwise? The design of the law, economic incentives and standards are all important here."

2. What is the most effective approach to implementation to ensure the success of Open Communications in enabling innovation and delivering the best consumer outcomes?

2.1. Irene Ng believes that giving data subjects access is the most effective approach – however not under consent provisions but as subject access request. In doing so, both

the individual and the firm benefits. The source firm can absolve itself from liability from the user re-sharing to the SAR agent or TPPs and the user gets their own data free from encumbrances. However, as stated in 1.2. the current law for GDPR does not enable subject access request through an API.

2.2. Irene thinks the current market for data sharing is not broken. For example, Amazon Alexa can request a user's Spotify playlist with the user's permission. Where companies enable API access to the data held by them, their benefit is one of lock-in. Any third-party application that takes on the data creates a dependency on the source and the exchange is mutually beneficial. The source gets greater lock-in benefits and the destination gets to provide a better service from the data.

2.3. Where companies do not enable API access, it may be because they operate legacy systems that cost too much to change and the firm does not see the benefits from updating the systems. Yet, this is efficient as well, because data that cannot be shared can also not be easily updated. Such data would be static, non-representative and less competitive. In the language of the consultation document, data that is not smart quickly becomes stale and non-competitive.

2.4. One solution to matching subject access to Smart Data with the maturity of the firm is to link explicitly to the sharing incentives of the firm. If the firm shares Smart Data with others, for example Fitbit granting API access to MyFitnessPal, to gain benefits for their business, then the firm must be obligated to share Smart Data with the customer through an API-enabled subject access request. If the firm does not share Smart Data with any other company, then it is not obligated to do so with their own customers. This ensures the market has a role in creating more efficient outcomes.

3. Are there any further actions we should take to enable consumers to benefit from Smart Data in regulated markets?

3.1. As outlined in Section 1, both Jon Crowcroft and Irene Ng highlight the importance of considering different models of ownership for aggregated data, i.e. centralised vs federated vs de-centralised models. There should be a legal and economic differentiation between treatment of these. However, current law for centralised and decentralised data only applies in regulated markets, for example cryptocurrencies. There is a benefit in bringing the centralised, federated and decentralised systems closer together, as they often operate with the same underpinning technology and are intended for similar outcomes.

3.2. Melanie Smallman adds that alternative models of company ownership, wider diversity in the tech workforce and more citizen participation in innovation decision-

making all have the potential to benefit consumers, particularly in helping innovators recognise that different social groups face different issues that they need new products and services to tackle.

3.3. It is worth considering the "switching costs" of switching service providers in greater detail. Anya Skatova believes that we should not automatically assume that consumers will prefer to switch providers in all circumstances based on apparent monetary benefits. In some markets, such as the energy market, many consumers appear to leave 'money on the table' when offered the chance to buy a homogeneous product at a lower price[3].

4. In which other markets, outside of the regulated and digital markets, would there be the greatest benefits from Smart Data initiatives? Please explain your reasoning.

4.1. Irene Ng advises that all markets would benefit from such an initiative, but that it's important to remember that data is a systemic resource i.e. it works "vertically" within each sector, and "horizontally" across user contexts that cross sectoral boundaries and needs to be managed accordingly. Managing data as a systemic resource can mean that Smart Data can be used for better coordination in times of crisis, as a resource for system-level outcomes, for example in combatting inequality and poverty, and used a store of value and medium of exchange across sectors. She believes that the pre-requisite for this would be the legal and economic separation of consent and GDPR rights from data subject access and control, as elaborated in 1.2 above.

4.2. Smart Data initiatives, with the underpinning data infrastructures, can involve considerable financial and reputational risk. Some of the possible lower-risk approaches or starting points, highlighted by Turing researchers, include:

- Academic research, for example citizen "Data Donation" programmes[4]
- Personal transport data can be used to encourage modal shift and meet environmental / congestion targets[5]
- Medical test data such as x-rays, CT scans, and other diagnostic imaging – for training / machine learning from medical test data

[3] https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3024534

[4] https://psyarxiv.com/ab83q/

[5] Proposed/accelerated in Data Pitch (H2020 Grant Agreement 732506): https://cordis.europa.eu/project/rcn/206193/factsheet/en

- Electric vehicle data, for example home charging data, garage/retailer/service station charging data, EV data from manufacturers – for innovation in EV infrastructure
- Food retail data, for example basket information – for farm-to-market innovation[6]

Philip Treleaven highlights the potential of creating digital marketplaces in the professional services markets such as audit, accountancy and tax, to help collaborators easily exchange documents, serve clients more efficiently and promote innovation by client companies supporting and funding new start-ups.

4.3. In line with the systemic nature of data as a resource highlighted in 4.1, Peter Grindrod highlights that there are three disjointed spheres that exploit customer / user data: academic, public sector, and commercial. These are governed by different forces, and while the people employed in these spheres may apply similar methods to similar datasets, we should avoid imposing the restraints and rules of one sphere onto another. Confusion arises when institutions switch spheres or when ethical codes, standards and practices developed for one sphere are applied to another, for example when a tech company conducts research without academic ethical approval yet seeking to publish in academic journals[7].

5. What other roles might industry find it useful for government to perform in addition to acting as a facilitator for Smart Data?

5.1. Several Turing researchers agree that the Government could lead the charge. Irene Ng believes that citizen data should be the first to be granted subject access through APIs. Jon Crowcroft advises that the Government can set good examples of being transparent about the reasons certain types of data can or cannot be made available, for example tax, education, healthcare, Border Force records.

5.2. Incentives should be put in place to discourage the acquiring of data, and incentivising the inquiring of it. This would help reduce cyber security risks and prevent copies of the same data held in multiple places, which can lead to fragmentation and legal hoovering of personal data. While adherence to GDPR is a disincentive already, we would welcome greater action in this area.

Jonathan Cave also highlights the importance of being a reliable "facilitator." In the public administration example, the "Once Only" principle at the European level is a

[6] Learnings from QROWD project (H2020 Grant Agreement 732194) http://qrowd-project.eu/project/

[7] https://royalsocietypublishing.org/doi/abs/10.1098/rsta.2016.0132

good first step to reduce the administrative load on individuals and / or businesses and address the risk of data fragmentation.

6. Do you agree that we should establish a cross-sector Smart Data Function with the proposed responsibilities set out above?

6.1. Establishing a cross-sector Smart Data function is not necessarily the right approach. An alternative way forward can be the refinement and/or extension of existing regulatory bodies can be an alternative approach, for example the FCA is responsible for open banking, the ICO for data breaches, Ofcom for open communications conformance, etc.

6.2. Irene Ng says unless the law is designed to separate "Smart Data subject access requests" from consent, and offer better protection for the former, a new cross-sector data function may not be effective. She adds that the regulated markets already have a Smart Data function: the FCA is already implementing data conduct, in the way it registers Account Information Service Providers, i.e. account information is the equivalent of Smart Data in the banking sector. She believes there is no evidence that the Government should intervene on Smart Data function across sectors – unless it is specifically for data subject access rights

7. What would be the best form for the Smart Data function to take? Should it be, for example, a new body, part of an existing body or some other form?

7.1. Florian Ostmann highlights several practical considerations. Firstly, proposals should consider the capacities and organisational knowledge of existing government departments and regulatory bodies and be mindful of the challenges associated with cultivating the relevant capacities and knowledge in a new body. Secondly, they should consider the importance of having a clearly identified locus of responsibility for Smart Data initiatives and of the institutional ability of the relevant body (or bodies) to effectively pursue the intended goals in the context of other data-related policy objectives pursued by different parts of government. Finally, the assignment of responsibility should carefully consider to what extent the intended benefits of 'smart data' can be achieved through measures and interventions that are confined to individual sectors or industries and to what extent their realisation requires cross-sectoral initiatives and coordination.

7.2. Elena Simperl advises that the new function must be well-funded and have enough resource to prevent it from impeding data innovation, and potentially leading to work-arounds that supersede it. The new function should have strong in-house expertise in GDPR, data sharing and data innovation, and should ideally be ring-fenced so that

resource competition and competing interests do not affect it. Elena recommends looking at two projects as examples of setting up similar projects: DataPitch.eu, an open innovation programme bringing together start-ups and corporate data owners, and SmartCityInnovation.eu, a similar programme where cities reach out to start-ups to make their data 'smarter.'

8. How can we ensure that the costs of Smart Data initiatives are shared fairly between the participating businesses?

8.1. Jon Crowcroft believes that data sharing is already viewed as a high-value proposition for businesses, so this is not a prevalent concern. The barriers to sharing are often due to legacy system use and would depend on updating technology. Where there is strong cultural resistance, political will should be applied.

8.2. According to Irene Ng, Smart Data initiatives should be setting the standard of what type of sharing is to be incentivised and what type of sharing is to be prevented due to greater risks and potential abuse. If the right incentives and economic payoffs are in place, firms can invest in being more mature in their data practices. The economic gains would be for setting, certifying and compliance to standards. She thinks the British Standards Institute is the right body to consult with on this question.

Using data and technology to help vulnerable consumers

9. What other actions could the government or regulators take to support the use of data and innovative services to improve outcomes for vulnerable consumers?

9.1. Bonnie Buchanan advises that in any final documentation, a working definition of "vulnerable consumer" needs to be provided. For example, financial vulnerability is related to low financial resilience, and could include experiences such as low financial capability, a divorce, a bereavement, or physical / mental health issues that affect one's daily activities in a significant manner.

9.2. Elena Simperl agrees that there are different kinds of vulnerability that need to be explored at a deeper level. The economically vulnerable might benefit from a more functional market – especially if they are time-poor and penalised by not switching providers. However, there are many vulnerabilities that will be exacerbated by data collection / sharing and the inherent risks.

9.3. Melanie Smallman agrees that the benefits of data technologies won't be spread fairly and evenly across society and early signs are that they are likely to reinforce existing

inequalities, i.e. the lowest income households will experience the most downsides and fewest benefits.

9.4. Irene Ng recommends a practical step to introduce regulation that ensures no application has the right to deny an individual their data, including inferential data, for example credit ratings. This can only be possible if the data within firms can be retrieved through subject access rights, and data can be retrieved by a guardian if the consumer is a minor, or his estate if the consumer is deceased.

10. Should we strengthen the powers of sector regulators to enable them to use consumer data to improve their understanding of the challenges faced by vulnerable consumers and to intervene to improve outcomes?

10.1. Irene Ng agrees that strengthening the sector regulator powers would be beneficial. Rights to access and use of data, if permission is given, must be only for a specific type of processing / usage within a specific timeframe. The regulators can provide mechanisms to enforce this.

10.2. Jon Crowcroft advises that the ICO has set some good examples recently, and the sector regulators should support these[8].

10.3. Bonnie Buchanan adds some additional considerations: "Considering the ACLU Idaho case, if carers or vulnerable customers are denied services, or lose benefits, will they be able to receive an explanation as to the data and model used in the decision-making process? Is there a provision for the vulnerable customer to regularly revise and update who their designated carer / attorney is? If an app can alert a trusted friend when a vulnerable person is in financial difficulties, how do regulators make sure this does not become more like surveillance?"

10.4. Melanie Smallman and Anya Skatova both highlight that we need to be aware of the effect of secondary use of data and overlaying data sets – for example, allowing police access to benefit claimant or health data, or overlaying health data with crime data, has the potential of subjecting particular individuals or communities to unacceptable levels of surveillance or suspicion.

10.5. Finally, Peter Grindrod brings up the example of biometrics as a case of personal data abuse in security and policing. He argues that it is important to make the distinction between human unique random biomarkers, for example fingerprints, and unique non-random biometrics, for example collected DNA samples and face images. Shared

[8] https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/consent/when-is-consent-appropriate/

ancestry by descent engenders practical ethical questions for police / security forces and regulators that need addressing. He believes more public debate and increased rigour in governance is necessary around adoption of biometric familial searching[9].

11. How can we ensure that the Smart Data Function improves outcomes for vulnerable consumers? Do we need to consider any further actions?

11.1. Jon Crowcroft believes that poor usability and high complexity can be a problem for consumers. While competitive offers are generally good, for example for annual renewal of car / home insurance, the cognitive overload associated with switching providers can be a challenge. Having proxy services that can act on behalf of vulnerable, for example elderly, young people, would be a good outcome, placing the cost on provider rather than the consumer.

11.2. Bonnie Buchanan would welcome a wider-ranging set of example services in the report. She talks about the case of the unbanked, who constitute 3% (1.23 million) of UK adults[10]. She believes the vision document does not go far enough in detailing how those without access to the internet or a smartphone may benefit from the Smart Data initiative. Additionally, as more bank branches close, especially in rural areas, how can Smart Data initiatives respond to the customers at risk?

11.3. Only 31% of UK adults perceive traditional financial institutions as honest and transparent[11]. Only 16% of UK adults rate themselves as highly financially knowledgeable. We think there is a long way to go in establishing more widespread trust in the financial system. There is a clear need for a Smart Data initiative in both data and financial literacy skills.

11.4. Anya Skatova advises that people value different types of data differently. Her research shows that people are more willing to pay for protecting financial and medical records than energy bills, loyalty card data or communications data.[12] They may have preferences over what is appropriate mode of use for each of the data types, and which level of protection should be used. She warns that public opinion polls are not always a reliable compass, as these preferences are often explained by lack of awareness about how their data is used, leaving people vulnerable to exploitation.

---

[9] https://www.researchgate.net/publication/333982366_Kinship_Familial_Searching_and_Biometrics

[10] https://www.fca.org.uk/publication/research/financial-lives-consumers-across-uk.pdf

[11] https://www.fca.org.uk/publication/research/financial-lives-consumers-across-uk.pdf

[12] https://psyarxiv.com/ahwe4/

11.5. Bonnie Buchanan believes that it would be interesting to see how the proposed initiative addresses the poverty premium issue in the UK – the idea that the poor pay more for essential goods and services and miss out of discounts or savings. The average cost of the poverty premium in the UK is £490.

Protecting consumers and their data

12. Do you agree these protections for when TPPs use Smart Data are needed? Are there others we should consider?

12.1. We think protections may not be effective if fundamentals are not sufficiently scoped. As previously mentioned, the proposed reliance on consent may not be the most effective approach. The researchers further challenge the emphasis on consent in this section.

12.2. On the topic of consent, Elena Simperl comments that consent messages are challenging, and there needs to be a robust process for creating and revising them. There is some prior work on how the success of a consent mechanism could be measured and scored, but it's mostly early stage. She points out that "explicit" is likely the wrong qualifier for consent. Consent needs to be unquestionable, but it may be entirely clear and more user friendly if it is contextual rather than framed explicitly as being "consent". The GDPR choice of "unambiguous" as an alternative qualifier is more appropriate.

12.3. Anya Skatova agrees that the topic of consent needs further exploration. When people don't fully understand the consequences of sharing their data, if benefits are overstated and/or risks are under-stated, there is a risk of exploitation. Additionally, we would welcome greater consideration on the question of obtaining consent in instances where multiple datasets are linked.

12.4. Elena Simperl explains while it's important to think through "approved purposes" of data sharing, there should still be some flexibility built in to allow for innovation. The converse of this is that in order to be meaningful to consumers, especially given the proposed reliance on consent, the purpose of data sharing needs to be described to consumers in specific terms.

12.5. Jonathan Cave believes complete contracts can be expensive and somewhat infeasible. There will be many instances where the interests of the parties evolve over time, highlighting the fact that meaningful consent needs monitoring and renewal, which may lead to 'consent fatigue'. Moreover, sometimes compliance is in neither party's interest – in 'efficient breaches' data sharing benefits both parties. He encourages thinking about networks of relationships, suggesting that reliance on

binary contractual relations may have significant limitations. At the end of the day, commoditised services will often require a degree of trust from the consumer: trust to believe what the businesses reveal; trust to act on one's behalf; trust that one's interests won't be compromised.

12.6. Elena Simperl notes that the discussion of UK's data protection legislation (Point 55 in the report) is like the "privacy seals" or industry co-regulation mechanisms envisaged in GDPR and should align to GDPR. She also raises some questions around the approach in Point 56:

"Is the implication that any other legal bases that might exist could not be applied to data received via a Smart Data initiative? Would the businesses' usual right to process data lawfully be restricted in this case?"

13. How should our proposed approach to accreditation operate in practice if it is to effectively ensure that consumers' data are protected and minimize burdens for TPPs?

13.1. Elena Simperl believes that the answer depends on the requirements for the accredited and non-accredited TPPs. She thinks there should be a basic threshold even the non-accredited TPP needs to meet. Perhaps the approach should be closer to 'accreditation' and 'higher accreditation.' Alternatively, there can be a technical solution that allows non-accredited / potential TPPs to use synthetic data for innovation purposes to create prototypes, and once the concept is proven they can be allowed to use it on real consumer data.

13.2. Jon Crowcroft adds that once streamlined, the FCA model could be generalised to other TPP accreditation areas.

13.3. Irene Ng warns that accreditation could disadvantage the accredited because TPPs would be operating within a narrow UK-level jurisdiction, whilst other companies could operate more widely.

14. What are the advantages and risks of introducing a cross-sectoral general authorisation regime for TPPs?

14.1. There is a risk that this regime may be harmful to both service providers and consumers. Jon Crowcroft suggests that trying several approaches and seeing what works best could be the best way forward, starting with areas with lower risks.

15. What other options should we consider to ensure that consumers are protected when using TPPs?

15.1. Please see the discussion on data infrastructure above. There is a range of privacy enhancing technologies (PET) that don't require TPPs to 'see' consumer data. A personal cloud is a viable alternative.

turing.ac.uk
@turinginst