Appendix T: our approach to assessing data remedies

1. Many of the potential interventions that we are considering in this study relate to data. This reflects the fundamental role that data plays in the business models of online platforms and publishers, particularly those funded by digital advertising. It is also consistent with the findings of the Furman Review, which advocated data mobility, interoperability and data access interventions.

2. There are specific considerations that need to be taken into account in to the assessment of data-related interventions. This appendix sets out some of these considerations. The objective is to explain some of the factors that have guided our assessment of interventions in the appendices and to explain the links between them.

3. The rest of this appendix is structured as follows:

   - First, we explain some of the important economic characteristics of data.
   - Second, we present a typology of the data-related interventions we are assessing in this market study.
   - Third, we set out several factors to be taken into account when appraising data-related interventions, including efficiency, privacy and competition.
   - Finally, we briefly summarise how these factors might be applied to remedy design.

Economic characteristics of data

4. Data has a number of distinctive economic characteristics which affect the incentives on platforms, consumers, and other market participants in digital markets.

5. At a high level, data has value because it generally enables people to make better decisions. This value is realised when people can collect, process, and use it.

6. Data typically has some characteristics of a public good.[1,2] Data and its derivative information is non-rivalrous (which means that it can be 'consumed'

[1] In economics, a public good is defined as being non-rivalrous and non-excludable. An example is the provision of national defence.
[2] See Stiglitz, J. E. (1999). Knowledge as a global public good. *Global public goods*, *1*(9), 308-326, available here; and Varian, H. R. (1999). *Markets for information goods* (Vol. 99). Institute for Monetary and Economic Studies, Bank of Japan, available here.

or copied infinitely without being ‘used up’). However, the extent to which data is also non-excludable (which means that it is difficult or infeasible to exclude someone from benefiting from data) depends on the technical and legal barriers to accessing and appropriating the returns to data.

7. In the context of online platforms and digital advertising, the relevant data is often highly excludable. Data controllers can often choose whether to share it. The excludability of data is borne out by the common complaints that we have heard from many market participants about their relative inability to access data to compete with Google and Facebook. The ability to control access to data is why firms’ own first-party sources of data are valuable. Data controlled by firms cannot be copied or accessed by others unless the data controller chooses to make it available, or there is some regulatory intervention to require access.

8. In the markets that we have studied, it is apparent that access to and use of data can be an important source of competitive advantage for firms in developing user-facing services and providing effective advertising in the form of personalised ads that fund these services. Differential access to data has implications for competition, market power, and the structure and concentration of these markets.

9. Whilst data itself is non-rivalrous, the returns and the competitive advantages from exploiting data (the information rents or innovation rents, depending on context) are rivalrous, and firms often have a strong incentive to prevent these rents from being competed away. In general, they can do this by excluding others’ access to data, using a variety of means such as maintaining secrecy, asserting intellectual property rights, or via technical measures such as degrading interoperability of data systems and frustrating efforts of others to reverse engineer or independently collect the data.

10. However, since data is non-rivalrous, exclusion may be inefficient as there could be large social gains if the data were widely shared and available. This insight underpins many general arguments for the importance of the free flow of data, which can create value for others and for society. The sharing of data can create significant efficiencies, where it allows data to be reused or combined in different ways and for different purposes. For the same reason, any intervention that fully restricts data sharing could be harmful for efficiency.

11. Typically, the marginal costs of copying and transmitting data are very low. However, there can be very substantial costs to create or collect it in the first place. In the context of digital advertising, these include costs to develop and innovate on providing valuable content and user-facing services that provide both inventory (opportunities to show ads) and insights about users (which

could be used to improve services and target ads), and also the costs of the technologies for collecting, organising, and sharing data about users at a large-scale. If firms cannot secure sufficient return on these activities, there may be too little investment in them relative to what is socially optimal. This trade-off between sharing data and providing adequate rewards to investing in the creation and collection of data is a familiar one from the extensive literature on optimal intellectual property rights.

12. Low marginal costs of copying and high fixed costs in creation give rise to economies of scale in the supply of data. In addition, there are often economies of scale and scope in the use of data to create insights. However, data is not homogeneous, but diverse in content and nature, and its usefulness and value depend on the type of data, its attributes such as freshness and velocity and the use to which it is put. Larger and different types of data, other things being equal, tend to be of greater value, in terms of both the insights that they provide generally and also the inferences they can support about specific individuals.[3]

13. Furthermore, data and certain types of data in particular can have multiple uses, not all of which can be foreseen by data subjects and data controllers when initial decisions are made about whether to use a service or to provide data, particularly in light of the rate of advances in technologies for data analysis and machine learning.[4]

14. These cost characteristics (low marginal costs of sharing, economies of scale and scope in the use of data) also mean that there is a value in aggregating data, either to make use of directly or to sell on to others, and a tendency to concentration. In many digital markets we have seen data aggregators emerging as a way of enabling data sharing.

15. In addition to considerations of efficiency and competition, much of the data that is relevant to online platforms and digital advertising is personal data - data which is about individual people. These individuals may receive different treatment based on this data and what is known about them, including personalised experiences and ads, personalised prices, or different access to opportunities and information. For this reason, and potentially many other motivations, many individuals have preferences for privacy and to be in

---

[3] It is likely that there are diminishing returns to scale over some range – for instance, it may be that after reaching a certain scale, adding more and more observations of a particular type of data may not be expected to yield the same value of insight that the first observations would. But conversely, it may be possible that certain techniques and models, and commercially viable levels of predictive accuracy and model performance, may only be obtained after amassing a very large scale of data. In other words, the 'production function' using data as an input may not be strictly concave. It may be concave in some regions but convex in others.
[4] See Appendix F.

control of information about themselves,[5] and individuals in the UK have rights under relevant data protection laws. Also, whilst many 'free' online services are provided partly in exchange for consumers' data for personalised advertising, consumers may find it relatively more difficult to assess the associated privacy costs than if they were being charged an explicit monetary price.

16. Designing policy and interventions to balance all these considerations and interests around the access to and use of data is a complex endeavour and an area of active research.[6] The rest of this appendix sets out our thinking on how to apply these considerations to the development of remedies in search, social media and digital advertising.

Typology of interventions

17. We have considered a range of different types of data-related remedies in this market study, reflecting the different ways data is used by the platforms and the range of competition and consumer issues that can arise.

18. The Furman Review drew a distinction between data mobility, interoperability, and third-party access. In comparison with the Furman Review, we have viewed data mobility as one way of increasing consumer control over data, alongside other potential interventions such as facilitating more informed choice to address the concerns raised in Chapter 4. We have also considered the possibility of restricting data flows within the platforms' ecosystems as an alternative to third-party access in some circumstances. And we have considered the potential for regulatory scrutiny or audit by a third party, particularly to address some of our transparency concerns.

19. Box T.1 sets out the main types of data-related interventions we have considered in the context of this study, giving some examples of each and references to where they are discussed in the report.

[5] DeCew, Judith, 'Privacy', *The Stanford Encyclopedia of Philosophy* (Spring 2018 Edition), Edward N. Zalta (ed.), available here.
[6] For instance, a recent paper by Jones and Tonetti (2020) shows, using a theoretical model with plausible assumptions, that if firms have de facto control rights over the data about consumers generated from their use of the firms' services, the outcome is that individual firms will collect and use as much data as possible, placing insufficient weight or ignoring the privacy preferences of their users, whilst restricting socially beneficial sharing of this data with other firms. Banning the sharing of data between firms would maximise privacy gains, but also result in a substantial loss in welfare as potentially valuable uses of data by others is left unexploited.

Box T.1: Typology of data-related remedies

1. **Increasing consumer control over the use of data**
   - Requiring consent for use of data (eg the intervention discussed in Appendix X that would require consumers to be given a choice over whether to receive personalised advertising).
   - Facilitating informed choice (such as the Fairness by Design duty assessed in Appendix Y, governing the use of choice architecture by SMS platforms).
   - Facilitating data mobility (eg through the use of PIMS, as discussed in Appendix Z in the context of digital advertising markets).
2. **Mandating interoperability**
   - Interoperability remedies in relation to social media functionality, discussed in Appendix W.
   - Restrictions on API deprecations through the Code, as assessed in Appendix U.
   - Potential standard setting interventions in digital advertising such as digital IDs and transaction ID, which are considered in Appendix Z.
3. **Mandating third-party access to data**
   - Click-and-query data remedy to overcome barriers to entry and expansion in search, as discussed in Appendix V.
   - Potential data access and transparency remedies in digital advertising, assessed in Appendix Z.
4. **Mandating data separation / data silos / restrictions on certain uses or sharing of data**
   - Potential interventions to address data advantages of large platforms through the creation of data silos, considered in Appendix Z.
5. **Allowing regulatory scrutiny and audit**
   - Applications across the issues discussed in Chapters 3 to 5, in particular through the Code (eg transparency of ad tech fees, regulatory scrutiny of auctions), as considered in Appendix U.

Factors to be taken into account in assessing potential data interventions

20. In this section we set out the broad factors that we have considered in assessing potential data-related interventions. This is not intended to be an exhaustive list, but rather to highlight the factors that are particularly relevant to our assessment across a range of different possible interventions.[7]

21. We have grouped these factors under the headings of:

- efficiency (static and dynamic);
- privacy; and
- competition.

22. There are often strong synergies between these effects – for example, between competition and efficiency – but in some cases there may also be tensions. The balance of these considerations may vary on a case by case basis, taking into account specific characteristics of the data and market context.

Efficiency

Static efficiency

23. As noted above, the economic characteristics of data mean that allowing it to be shared could lead to significant static efficiency benefits. This is because data can often be shared at very low marginal cost and different types of data can be combined to generate valuable new insights.

24. As set out in Chapter 3 and Appendix F, data is valuable both for consumer-facing services and in digital advertising. For example, on Facebook and other social media platforms, data on consumers' interests and characteristics can be used to decide what content to show and how it should be organised. Similar data can also be used to target advertising at consumers for whom it is more likely to be relevant and help assess its effectiveness. To the extent that this data could be shared, it could allow other competitors to take advantage of the same efficiencies.

[7] For example, we do not discuss here potential impacts on online harms or other wider impacts, which may be important in some cases.

25. We would expect the static efficiency benefits of data sharing to be greater if there are economies of scale and economies of scope in the use of data, and where there are positive externalities in the use of data. For example, social media platforms benefit from direct network effects through consumers being able to access others' content. Where this data is retained only within a platform's 'walled garden' and cannot be accessed by other platforms, the network effect works only to the benefit of that platform. By contrast, wider interoperability and sharing of content would allow network effects to be shared by other platforms in the market.

26. In assessing potential interventions, it is important to assess how valuable a particular type of data is in the provision of the consumer-facing services. In assessing this value, the following questions are likely to be relevant:

- What is the incremental improvement to the quality or viability of a service (or potential service), relative to a counterfactual without or with reduced access to this data? For example, how valuable is this data in delivering a strong internet search function? How valuable is this data in delivering effective social network?

- To what extent could changing the availability or access to this data overcome barriers to entry and expansion in the consumer-facing markets? Are there potential uses of the data that are currently unexploited and how valuable are these uses?

- Does the value or relevance of the data depreciate quickly? How valuable is access to historical data, relatively to access to near real-time data?

- Does acquiring data about one user or customer primarily lead to improvements in the quality of the service for that user, or to improvements that benefit all users of that service?

- Are there substitutes for this data (ie different data that could yield similar insights or information), and how available are these substitutes?

- What are the relevant costs of collecting or creating this data? How excludable is the data once it has been collected? Can the data be acquired by alternative means or reverse engineered[8] (for example from observing the outputs of the service)?

[8] The concept of Adversarial Interoperability is relevant in this respect.

27. Similarly, in relation to interventions relating to data for digital advertising, it would be important to assess how valuable the data is, asking similar questions to those posed above for consumer-facing services. For example:

- What are the estimates of the incremental value to advertisers, publishers, and adtech providers from having this data (for targeting, for attribution, and for other activities such as fraud prevention), relative to a counterfactual without or with reduced access to this data (but including other potentially substitute data that would be available anyway)? To what extent would the data enhance the effectiveness of advertising, as considered in Appendix O?

- To what extent could liberalising access to this data overcome barriers to entry and expansion in the digital advertising markets?

28. We would also need to factor in the costs of opening up access (in addition to possible dynamic efficiency impacts, as discussed below). These include practicalities and transactions costs. Relevant questions could include:

- Would there be any risks or costs involved in collecting / sharing the data securely? Who should bear these risks and costs?

- What is the impact on hassle costs for the consumer? To what extent would it be reasonable to expect consumers to make informed choices about the processing of their data in the relevant context, or to bear the burden and responsibilities for making choices that they are not well prepared to make?

29. Overall, the static efficiency benefits of third-party access are likely to be greater (and costs of data silos greater) where data is valuable in overcoming barriers to entry and expansion, and where there are economies of scale and scope in the use of the data. On the other hand, the case for greater consumer control may be weaker where this imposes excessive transactions costs, or unreasonable expectations, on the consumer and where there are strong externalities in the use of the data.

Dynamic efficiency

30. In addition to static efficiency effects, there are also important dynamic considerations which need to be taken into account in assessing the long-run efficiency impacts of any intervention. There can be very substantial costs to creating or collecting data in the first place. In the context of digital advertising, these include costs to develop and innovate on providing valuable content and user-facing services that provide both inventory (opportunities to

show ads) and insights about users (which could be used to improve services and target ads), and also the costs of the technologies for collecting, analysing, and sharing data about users at a large-scale. If firms cannot secure a sufficient return on either or both of these activities, and others are able to free ride on their innovations, there may be too little investment in them relative to the social optimum.

31. Conversely, in other contexts, data may have been gathered largely as a result of self-reinforcing incumbency advantages, or even through the leveraging of market power to expropriate others' data. In such cases, the interests of innovation may be met by freeing up access to the data, or imposing restrictions on its use by the incumbent.

32. In assessing the impact of specific interventions on dynamic efficiency, it is helpful to distinguish different cases on the basis of how the data has been collected. For example, relevant questions would include:

- Is this volunteered or observed user data, or is it inferred user data?[9] It is more likely to be the case that volunteered and observed data are 'raw' facts that require relatively low investment and effort to collect or create, whereas 'inferred' data may be the result of processing, combining and analysing data and may require more investment to create.[10] If so, the sharing of 'inferred' user data may be more likely to damage dynamic incentives by undermining the added value that has been created by a particular data controller; by contrast, the sharing of raw data might have less impact on dynamic efficiency.[11]

- More generally, to what extent has investment been required and are costs incurred on an ongoing basis to create and collect this category of data? Related to this, to what extent is the data being collected a key feature of the business model or more of a by-product of operating a service? Where costs are larger, we would expect the potential negative dynamic efficiency impacts to be greater.

- Who incurs the cost of creating or collecting this data? Who controls any benefits that arise from using the data, and who receives these benefits?

---

[9] See Appendix F for a description of volunteered, observed and inferred data.
[10] The distinction between volunteered and observed data on the one hand, and inferred data on the other, is recognised within the right to data portability under GDPR, which allows individuals to obtain and reuse their personal data for their own purposes across different services. It only applies to information that an individual has provided (volunteered) or data resulting from observation of an individual's activities (observed). It does not include any additional personal data that has been created from volunteered or observed data. (ICO, Guide to the General Data Protection Regulation (GDPR), pp.128-129.)
[11] However, in some contexts, the costs of collecting 'raw' data can itself very costly, particularly if these costs include those of creating valuable consumer-facing services that attracts users and give rise to the opportunities to collect data in the first place.

Similarly, are there any risks or costs from various uses of the data? Do these costs and benefits accrue to the same entity, or are there externalities?

- Are the benefits from the collection and use of data being expropriated from those who bear the cost or risks from its use, due to an imbalance of negotiating position or exploitation of market power?

33. As with the consideration of static efficiency impacts, there are also practical issues and transaction costs that would need to be factored into any assessment. Among the relevant questions are:

- What are the challenges in creating a stable regulatory solution in markets that change rapidly over time?

- In relation specifically to mandated interoperability, is there a risk that the intervention will restrict innovation and diversity through excessive standardisation?

- Can an appropriate access regime be defined to mitigate potentially adverse effects, such as through charging for the use of data?

34. Overall, in relation to dynamic efficiency impacts, where data is currently held by a platform without significant ongoing costs, or where it has been gathered as a result of incumbency advantages, the case for third-party access and mandated interoperability is greater. Where there is innovation or investment required to collect data, the dynamic costs of greater third-party access or mandated interoperability are likely to be greater. These costs might be mitigated in some cases through defining an access regime, for example including suitable payment terms.

35. We take these factors into account in a range of the data-related interventions we consider in this study, and in particular the potential interventions to require Google to provide search click and query data to third parties and to require Facebook to increase interoperability in respect of certain social network functionalities. These are assessed in Appendices V and W respectively.

Privacy

36. Much of the data that is relevant to online platforms and digital advertising is personal data, ie data which is about individual people. These individuals may receive different treatment based on this data and what is known about them, including personalised experiences and ads, personalised prices, or different access to opportunities and information. For this reason, and potentially many

other motivations, many individuals have preferences for privacy and to be in control of information about themselves,[12] and individuals in the UK have rights under relevant data protection laws.

37. As a result, there may sometimes be tensions in some cases between the efficiency benefits of greater data sharing or interoperability as discussed above, and the privacy concerns of consumers who may not want their data to be shared. Equally, there may be legal and other constraints on the sharing of data resulting from data protection laws.

38. In principle, we believe that informed and motivated consumers and data subjects should be well-placed to balance their privacy concerns against the benefits from using their personal data.[13] For this reason, we have been very interested in exploring interventions that would enhance consumers' ability to access, assess and act on relevant information about the uses, purposes, and controls that they have over their personal data.

39. However, we also recognise that many people may reasonably not want to spend time and effort investigating the implications of different privacy policies, and may not therefore be well placed to understand how their data will be used, and to assess and evaluate abstract privacy risks against the benefits of accessing the service or information that they want. As discussed in Chapter 4 and Appendix L, this is exhibited in the low engagement that typical consumers have with privacy and cookie notices, and the phenomenon of 'consent notice fatigue' as consumers, almost reflexively, ignore or quickly click past such notices. For this reason, we are mindful of the impact on potential interventions on the burden on individual consumers to exercise their data protection rights effectively, and to bear the risks and costs potentially arising from the processing of their personal data.

40. Therefore, alongside exploring a number of measures that enhance consumer control over the use of their personal data (for example, in Appendix X), we place emphasis on approaches that can reduce friction and facilitate informed decision making on the part of consumers, notably through the Fairness by Design duty on platforms discussed in Appendix Y. We also recognise the limits of informed consent, and that it may sometimes be necessary for the

---

[12] DeCew, Judith, 'Privacy', *The Stanford Encyclopedia of Philosophy* (Spring 2018 Edition), Edward N. Zalta (ed.), available here.

[13] There will be exceptions to this general position, such as where there are substantial positive externalities from certain uses of personal data. In these cases, it may be justified to compel individuals to provide their personal data for these uses, such as the compulsory participation in national censuses. Similarly, there will be instances where there can be negative externalities from an individual's choices about the use of their personal data, such as situations where revealing information about one individual could also reveal information about others. See Floridi, L. (2014). Open data, data protection, and group privacy. *Philosophy & Technology*, *27*(1), 1-3, available here. See also Taylor, L., Floridi, L., & Van der Sloot, B. (Eds.). (2016). *Group privacy: New challenges of data technologies* (Vol. 126). Springer, available here.)

government (including regulators such the proposed Digital Markets Unit) to play a more interventionist role and directly determine whether certain uses of data should be permitted or restricted, and control over data rebalanced, taking into account the overall costs and benefits to society, including in terms of efficiency and individual consumers' rights to privacy and data protection.

41. The key questions that would be relevant to the assessment of data interventions in respect of privacy include:

- To what extent can the data be linked back to individuals, using reasonably available means (including other data sources)? Is it personal data, within the meaning of GDPR, and therefore subject to the rights and responsibilities of relevant data protection legislation?
- Is the data aggregated or anonymised? Can the data be anonymised without losing its value? What are the risks of re-identification or deanonymisation in the relevant context?
- Is it sensitive personal data (ie special category data)? What could go wrong in the event of a security breach in relation to this data? What protections are there to limit the risks and costs to consumers?
- Does the intervention place an excessive burden on consumers to make informed choices about the processing of their personal data, or to protect their privacy, that they may not be well positioned to make?
- How can genuinely informed consent be facilitated while minimising the burden on the consumer?

42. We note that a regulatory intervention (for example through order making powers or legislative requirements) could, in principle, provide a legitimate basis for processing. Therefore, the current parameters of GDPR and PECR do not necessarily create a decisive constraint on which interventions are possible. Because of this, it is important to continue to understand evolving consumer attitudes and capabilities; for example:

- Do consumers know that the data is collected and how it is used?
- Are the consumers concerned about the use to which the data is put? Are there concerns about security standards?
- Do consumers care about sharing the data? Within the company that collects it? With third parties?

- How much variability is there in concerns among consumers? Is it possible to identify a ‘representative’ consumer when determining the appropriate interventions?

- To what extent do consumers exercise any controls that are already available to them?

43. We consider the current evidence on consumer attitudes and behaviour in relation to data sharing and privacy in Chapter 4 and Appendix L.

Competition

44. Finally, we have assessed the impact of potential data interventions on competition. Competition is not an end in itself, but can be a powerful force in supporting efficiency and innovation, so that firms deliver better services for consumers over time and provide more of what consumers value (potentially including privacy protections) at lower cost. Competition also facilitates and in turn depends upon informed consumer choice, and so is strongly supportive of greater consumer control.

45. In our study we have identified several remedies that seek to promote both competition and privacy. For example, the choice requirement discussed in Appendix X would require platforms to give consumers more choice over the use of their personal data in respect of personalised advertising. It therefore aims to support privacy but also to help promote competition by changing the nature of the relationship between consumers and platforms, by giving consumers not simply more choice but more control, and putting the onus on platforms to engage consumers rather than relying on and encouraging consumer inertia.

46. There are also generally strong synergies between competition and increasing access to data (which can be efficient, due to the non-rivalrous nature of data, as outlined above). For example, overcoming barriers to entry and expansion resulting from the scale and scope advantages of the large platforms’ access and control of data can increase competition.

47. However, we note that it is also possible for there to be strong competition with relatively limited data sharing. One example would be an intervention to limit the degree of data sharing within large platforms’ ecosystems (‘data silos’), discussed in Appendix Z. This might reduce the static efficiency benefits of data sharing described above. However, it would also reduce the competitive advantage of the large platforms over other competitors with more limited first-party data sources, and so could have a positive impact on competition.

48. We think that an important overarching principle that can be applied in designing regulatory interventions to balance these privacy and efficiency considerations is competition 'neutrality'. A neutral regulatory system is one which treats functionally similar activities similarly and avoids favouring one kind of business model over another in a way that could damage competition or harm efficiency. Regulations that are not competitively neutral are prone to regulatory arbitrage, as firms organise their activities to circumvent regulatory burdens in costly ways.

49. In principle, competitive and well-functioning markets for online services and digital advertising are compatible with both efficient use of data and greater privacy protections. There can be trade-offs, however, between the latter two considerations. At a high level, an emphasis on efficiency may involve greater data sharing or interoperability, while an emphasis on privacy may involve restrictions on the platforms' use of data for different activities. In both cases, we would expect to see competition benefits as a result of a reduction in barriers to entry and expansion.

Application of these factors to remedy design

50. This section draws some summary conclusions about how the principles outlined above might translate into remedy design for the different types of interventions listed in Box T.1 above. The case for specific interventions is assessed in more detail in the appendices.

51. Overall, we think that the case for consumer control remedies is likely to be stronger where:

- Consumers are well-informed, confident and effective in their exercise of controls that are made available to them.
- There are no significant externalities, ie costs and benefits that affect others who are not part of the decisions between consumers and the data controller (such as a platform, or an online service provider) about whether and how to use that data.
- Consumers' views vary substantially (so a one-size-fits-all rule would not be appropriate).

52. The case for greater consumer control may be weaker where this imposes excessive transactions costs or unreasonable expectations on the consumer, and in the presence of broader costs and benefits to society.

53. The case for third-party data access and/or mandated interoperability is likely to be greater where:

- data is valuable in overcoming barriers to entry and expansion, and there are economies of scale and scope in the use of the data;
- data is held by a platform without significant ongoing costs or investment, and mainly due to historical first-mover advantages;
- privacy concerns can be effectively managed (eg where the data is not personal or can be effectively anonymised or aggregated).

54. The case for mandating data silos is likely to be greater where:

- privacy and competition concerns are high, but consumer control remedies are impractical (because of complexity, time, incentives, externalities, etc);
- data is gathered through leveraging market power to expropriate others' data.

55. Regulatory scrutiny and audit can help improve trust where direct access to data may not be appropriate due to competition, privacy or gaming concerns.