Research Article

Katharine Rockett*

Is Data the New Gold? Considering Intellectual Property Protection and Regulation of Data#

https://doi.org/10.1515/econ-2022-0040
received December 02, 2021; accepted April 13, 2023

Abstract: This article reviews the economics literature on the optimal design of intellectual property, the current framework for protecting data, focussing on the European Union, and the main features of data. It combines these to ask how current intellectual property protection instruments and surrounding regulation could be adapted to more usefully protect (non-personal) data that require investment to create and encourage its diffusion. The main changes from current protections include an expansion of registration and cataloguing systems, more definition of quality certification, and increased attention to portfolios of protection mechanisms.

Keywords: data, intellectual property protection, trade secrecy, registration

1 Introduction

A quick search on the Internet for the phrase "Data is the new gold" turns up hits from the World Economic Forum, Forbes, and Deloitte among many others.[1] It is an appealing image. The gold rushes of the nineteenth century were characterised by the dream of great wealth for those of all walks of life, even though it provided this to only a few. Property rights were at the heart of the behaviour of agents during the Gold Rush. Clay and Wright (2005) study the records of the US Gold Rush, beginning in 1848, finding that claims to specific property rights that were established soon after the initial rush resulted in a race to establish claims where speed was all important. These rights evolved quickly. Within a few months, it was accepted that if one left tools in a hole, that hole was not to be tampered with. Property extended to the idea of a claim as land, rather than a hole, within a year. At the same time, these rights were far from secure, with legal ambiguity, dispossession, and abandonment common in early years. The rights that evolved were also heavily influenced by earlier laws and adaptations to them. Notable among these were the Mexican mining law, which established rights based on discovery and development and "working," and the earlier Homestead Act, which established family-sized tracts based on squatting alongside work to improve the land.

What can we say from this starting point about property rights to and regulation of data and how these should be designed to get the most for society? The idea of rights that encourage rapid development of a sector is attractive, but at the same time, the abuses and problems caused by initial design problems for rights, as was seen in the Gold Rush, serves as a caution, particularly where rights cover unfamiliar territory.

Of course, despite the superficial similarities between data and gold in the feeling of a "rush," the significant differences between the two suggest that we cannot carry over property right lessons from gold to data. Gold is an exhaustible resource; data are not. Gold may have more intrinsic value than data. Gold cannot be used simultaneously by multiple agents; data can. These differences mean that the optimal protection likely will not be the same. At the same time, observing the differences between these two cases can serve as a tool to tease out the nature of appropriate property rights on data.

The purpose of this article is to look at data's features, consider the data rights landscape as it currently exists (mainly in Europe) and, starting from a base of the standard intellectual property rights (IPRs) literature, suggest adaptations to the existing landscape that might help to accommodate data's unique features. The paper does not

A preliminary version of this article was presented as a keynote talk to the MPI Alumni Conference, July 9, 2021. Many thanks to the audience at that conference, Suma Athreye, Nicola Searle, and two anonymous referees for their helpful comments.

1 See Deloitte (2018), Forbes Africa (2019), and Halloran and D'Souza (2020) as some of many examples available on Internet search.

* **Corresponding author: Katharine Rockett**, Department of Economics, University of Essex, Wivenhoe Park, Colchester, CO79LD, United Kingdom, e-mail: kerock@essex.ac.uk

address creating a new right for raw data (Kerber, 2016) and excludes data that take no investment to create such as data that are produced in the course of doing normal business. This digitisation of use does not require the standard incentives to create that are the basis for intellectual property protection's design. Instead, the article focuses on processed data, derivative data that must be created with effort, and other data requiring investment and current protections for these types of data. Processed data, in other words, data that have been manipulated to become easier to access, manage, analyse and update by means of investment, have been discussed recently in light of the modest uptake of and mixed feedback on the *sui generis* database protection. This right will serve as a focus and a point of departure.[2] The paper skirts other important issues such as protecting the sources of data, restrictions (mainly privacy concerns) on the curation and use of personal data, and transparency issues such as profiling.[3] The aim is not to present a comprehensive analysis or all the answers for all types of data, but rather a simple framework that raises some limited questions for rights design.

Other works have been done on this issue: the European Commission (EC) has published (EC, 2018) a set of consultation papers on the Database Directive European Parliament and Council (1996) that aim to evaluate pruning and adjustment (or, indeed, elimination) of the right.[4] WIPO (2002) summarises progress in various countries on protection of non-original data, presenting a menu of features for design and adjustment found in national legislation. Our article, in contrast, focuses on the conceptual arguments for IPR design found in the literature, rather than on the empirical aspects of the landscape. The literature review raises some areas where current rights and regulation could be adapted, focussing on registration of information, the design of rights and regulation where many protection measures are possible to use as a bundle, and how to incentivise agents to take up any data protection rights. One could think of this article as a response to the puzzle in the 2018 EC consultation paper on why database protection is not used more and possible modifications to this protection to improve its effectiveness and uptake. In contrast to the work of Graef and Prufer (2021), this article aims for a few straightforward changes. While recognising the validity of the analysis of the need for a new right in the study by Drexl et al. (2016), we find potential merit in modifications to existing database protection and trade secrecy rules as applied to data. One could also think of the paper as posing issues for further work for the European Data Innovation Board (EDIB), established by the Data Governance Act (as amended, 2022),[5] as this article takes a more interventionist approach to the quality and access to data than the Act.

This article will argue most strongly for a centralised registration system for (meta)data, rather than simply registering data intermediaries, as in the Data Governance Act, to further facilitate trading. One could think of this as a catalogue and index system for data. This registry would be the map, if you will, of the currently staked-out claims with a description of the claim but not revealing the precise content. In the same way as other IPR registration systems promote trade through allowing partners who are initially unknown to each other to interact, data licensing and competition in data pricing could be promoted by a trustworthy one-stop shop for information on the data landscape.[6] This central catalogue and index may be generated by the government or built via interconnection with the market data exchange market. The article also recommends that minimum quality certification (of content) be implemented directly or required from market traders. This is a provision already roughly addressed in the Data Governance Act (article 12, paragraph g), although the hope is that this article makes the aim of this quality certification more precise. A government-sponsored quality mark showing sufficient quality has the advantage of promoting trust in the market.

Finally, the article argues that uptake of the right will be challenging unless the "default" of other protection measures, in particular secrecy without revelation or and

2 Compared to raw data, processed data usually have some benefits including cleaning, curating and consistency, but the degree of benefit depends on the type of manipulation that has occurred. As Graef and Prufer (2021) note, processed data involve an investment and normally should be treated distinctly from raw data. See Graef and Prufer's (2021) summary of the recent EU legislation on raw data sharing and their argument for mandatory data sharing for raw user data in the European context. While the Data Governance Act applies quite widely to data, the focus of this article's literature review on intellectual property would make raw data out of scope. The more recent Data Act focuses on data produced in the course of normal business use, and also clarifies that the Database Directive does not apply to that type of data, one of the former confusions around that protection.

3 This is not to deny the enormous amount of work being undertaken in privacy research now. See Acquisti et al. (2016) and Goldfarb and Tucker (2019) or, for an interdisciplinary review, see Smith et al. (2011).

4 See also Reichman and Samuelson (1997) for an early consideration of the optimal design of data protection. This article considers changes to the current protection regime, and so is not as wide ranging as that piece.

5 See Data Governance Act, 2018/1724, as amended by Regulation 2022/868, 30 May 2022. European Parliament and Council (2022a).

6 See Graef and Prufer (2021) for a summary of EU raw data sharing protocols, including a register of traders.

protection with no end date. Many alternative protections are clearly attractive to data holders currently but are not necessarily the optimal choice for society. While these alternative measures are not fully analysed here, the point is that any analysis of database protection must be considered in tandem with its alternatives.[7] By and large, this is not the approach that the literature on intellectual property protection instruments has taken. This article argues that options that do not involve the revelation of the data and a time limit are inferior in this case and so takes a more interventionist approach to channelling the choice of data management and diffusion towards registration and potential trade. In terms of trade secrecy, this suggests a possible data exception. Again, this is hinted at in the Data Governance Act as one possible interpretation of Article 30, paragraph d, but the hope is to make it more precise here.

The article mentions in passing as part of a potential menu of issues to pursue as the framework for data governance evolves, the protection of independently invented data,[8] the desirability of lenient grantback provisions, and FRAND[9] licensing terms. These could be useful to address both sequential innovation concerns and concerns about pricing of complementary inputs to follow-on innovations that have been raised by the literature. A final consideration, raised by the Gold Rush, is whether one should only retain exclusive rights on data when the data are "worked" or whether purely defensive use of data should be permissible.

This piece takes a purely economic perspective on these issues. Clearly, however, there must be some overlap with legal scholarship if we are to examine the creation of a right. Two companion pieces in this volume consider, the same type of question from a legal standpoint, leaving this piece with the liberty of an admittedly narrower, economic, approach.

2 The Lay of the Land

The set of European legislation that impinges on data creation and use is as impressively complex as it is recent. Figure 1 lists selected legislation, including the newly announced Data Act.

The effect of this plethora of legislation is difficult to gauge, as some is still evolving. We do, however, have a 2017 study[10] relating to one piece of legislation, the Database Directive that generated *sui generis* database protection, which outlines many of the concerns that could be raised about the set of legislation in general.

That study shows that databases currently are protected using a variety of instruments, reproduced in Figure 2. These different forms of protection may be used together or separately to form the totality of protection for the data.

This feature of multiple instruments of protection has some similarity with earlier studies of patent rights, which have found that innovators may use patent rights alongside other types of strategic measures (such as first mover advantages in marketing or learning by doing advantages) or other legal measures (secrecy, for example) to protect their innovations.[11] The menu of available protection measures for data is not the same, however, as that earlier study of manufacturing innovations. For example, technological protection measures (e.g. technological "locks" to prevent copying or using the data without authorisation) were not relevant to that earlier survey. Strategic advantages such as learning by doing were investigated earlier but were not investigated in the 2018 survey, even though they might be relevant. In sum, while a range of protections may be relevant to data, and may be viewed as complementary or substitutable, it is not clear that

7 See argument below. Trade secrecy is not fully analysed here. Bessen (2005) and Cugno and Ottoz (2006) argue that trade secrecy dominates patents *ex post* innovation but do not show that this is true when *ex ante* incentives are considered. Anton and Yao (1994) show that strong trade secrecy protection can preserve innovation incentives, however, Fosfuri and Ronde (2004) find that trade secrecy can interact with the positive effects of aggregation economies under "job hopping" to generate benefits for all parties of weaker trade secrecy that accommodates spillovers. Their argument is not unlike Bessen and Maskin's (2009) argument against patent protection when innovations are complementary and where "frictions" exist in the market that generate profits from innovation for innovators in the absence of protection. Anton and Yao (2004) find that patents and trade secrets are selected as protection for different types of innovations where choice is available, and Risch (2007) finds that, where costless to provide, allowing trade secrecy dominates not allowing it.

8 Maurer and Scotchmer (2002) make an argument that independent invention benefits consumers and welfare by lowering prices (due to competition by close substitutes) without increasing duplicative research costs (because firms will select out of creating duplicate invention as prices fall). They note that narrow patents can also have a similar effect. Katznelson and Howells (2021) argue that narrow (but not too narrow) patents have the additional benefit of directing research optimally. This effect complements, and was not a focus of, the earlier work.

9 Fair, Reasonable, and Non-Discriminatory

10 European Commission: Study In Support of the Evaluation of Directive 96/9/EC on the Legal Protection of Databases Annex 2: Economic Analysis 2017/0084.

11 See the classic study by Cohen et al. (2000) for details.

Trade Secrecy Act 2017 – confidential information. While misappropriation restricted, reverse engineering and parallel innovation are not. Cease and desist orders, damages and injunctions possible.

Database Directive (revised provisions included in Data Act, 2023, below) – protects databases where substantial investment, even if not particularly creative work. (Creative work protected under Copyright).

General Data Protection Regulation (2016) – very broad legislation concerning personal data protection, curation, and flow.

Digital Markets Act (Active from 2023) – prohibits unfair use by data "gatekeepers" - durable, strong, platforms. Interoperability, data access, open access for customers to businesses outside intermediary, level playing field for products/services, allow uninstalling of software.

Digital Services Act (2022) – various hosts, intermediaries, and platforms. Requires transparency and compliance with fundamental rights and various codes of conduct.

Data Governance Act (2022) – improve trust through voluntary labelling, data-sharing structures including interoperability (esp. public sector data), horizontal board to oversee data governance.

Data Act (2023) – increase incentives and ability to share, create fair, and certain contracting and negotiation environment.

AI Act (2021) – protection of artificial intelligence. Related to but not directly protecting data.

Free Flow of Non-personal Data Regulation (2018), removing localisation requirements for data and self-regulating data porting including codes of conduct. Access for relevant authorities is upon request and to help those authorities fulfil their duties.

Open Data Directive (2019) on access to public sector information.

Figure 1: A non-exhaustive list of current laws affecting the status of data protection in the EU.

the relevant choices are the same as for patentable technologies.

Secondly, the protection afforded by the directive is unclear. Forty percent of the respondents to the survey indicated that they had some level of uncertainty about whether they had infringed the right because of poor explanation of the right and in view of the rapidly changing nature of databases. It may also be the result of poor awareness or the fact that individual member states have leeway in how they interpret the Directive's requirements. Of course, a right that is poorly understood or simply not known is not likely to be effective.[12] The recent multiplication of rights in Figure 1 may result in further confusion.

Finally, respondents indicated that the *sui generis* right did not necessarily increase their incentives to create new databases, amounting to less than 10% in all categories of innovation. As pointed out by the EC in its evaluation of the performance of the Directive, it was, "[i]ntroduced to stimulate the production of databases in Europe, [but] the instrument has no proven impact on the production of databases."[13] Reform of the right could be a better solution than eliminating it, however. Referring again to the earlier classic work by Cohen et al. (2000),[14] patents are not heavily relied upon as a means of protection on average across a wide swathe of sectors but rather are highly effective in a few concentrated portions of the economy such as pharmaceuticals and chemicals. If the same is true for data, then protection may well be worthwhile and socially quite beneficial even if on average it does not make much of a difference.

Overall, then, the current situation is confusing and evolving: a variety of protections may be applied to data, perhaps together and perhaps as substitutes, in different combinations across different industries and governed by a complex web of legislation. The lack of clarity may affect rights uptake. Designing a right that is useful and does no harm in such a landscape is the challenge to any proposed modification and aiming for clarity implies restricting attention to straightforward ideas, where possible. This does not necessarily mean light touch, but systems that are easy to understand, clear, and accessible would be the most helpful.

12 See Gans et al. (2007) and Heger and Hussinger (2017) for a discussion of the benefits of increased certainty in intellectual property rights for trade in intellectual property.

13 See Boyle (2006).

14 See also Dosi et al. (2006), Fontana et al. (2013), Moser (2012, 2013), and the references therein.

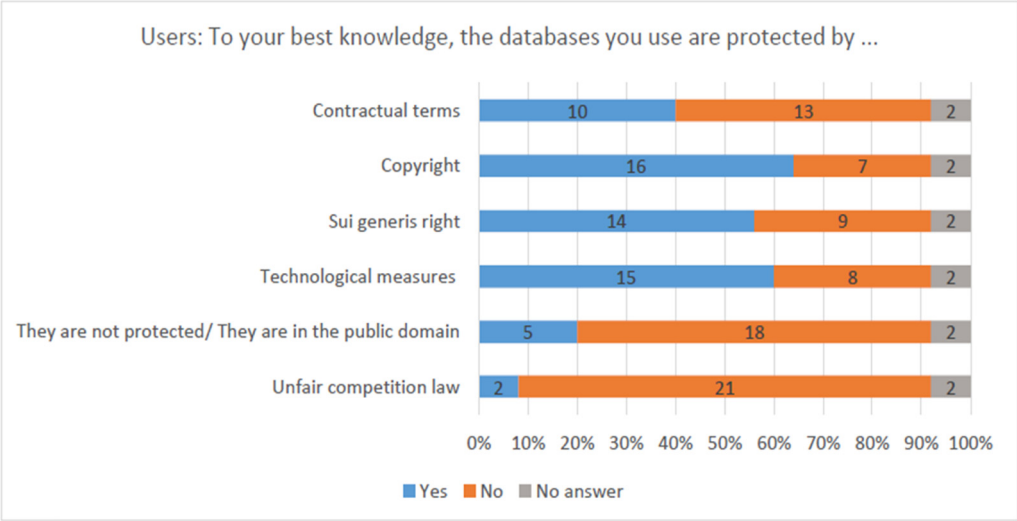

Figure 2: Reproduction of: European commission: Study in Support of the Evaluation of Directive 96/9/EC on the Legal Protection of Databases. Annex 2: Economic Analysis 2017/0084, Figure 15. "Means of protection of the databases by data users."

3 What is the Rush?

There is no fully functioning data collection system or protocol worldwide. Estimates are, however, available for the level and increase in data use. A recent EU report (European Commission, 2020) estimated that the value of the data market in the EU (EU27 plus United Kingdom) was 400Bbn euros in 2019, with a growth in value of 4.9% per year, exceeding the EU GDP growth by more than 3%. Revenue growth of data suppliers was also strong at 9% in the four years to 2019. Data professionals benefitted from this sectoral strength: the workforce of data professionals grew at 5.5% in 2018, with little sign of slowing. The "data as a service" portion of the market was projected in the same report to grow by 10% per annum in the 2021–2026 period.

These positive numbers for Europe are not unique. The same EU report also noted that the EU data market was around two and a half times smaller than its US equivalent. Statista estimates[15] that the global datasphere will grow to 51 zettabytes[16] by 2025. Of course, raw and personal data make up a proportion of this and are not really our focus here.

Still, these numbers suggest that this is hardly a market that is struggling to develop, so the argument for a change in protection is not entirely clear. If people are already rushing to the gold fields, do we really need to encourage more? Of course, robust growth does not necessarily mean socially optimal growth. To determine this, we would need to examine standard factors that drive a wedge between social and market optima. These include failures from strategic behaviour or lack of entry, externalities, and informational failures, as well as the current system of regulation and protections that may impinge on market functioning.

Considering these market failures in turn, several authors have raised concerns about strategic behaviour in data-sharing markets and the incentives not to share or to exclude entry, with strong network externalities a driving feature, although work in the area is nascent (Richter & Slowinski, 2019). Indeed, the threat of strategic behaviour has been recognised in the remit of the EDIB. In the context of personal data, Acemoglu et al. (2022) point to data's heterogeneity, with social value compared to private value likely to depend on the specific type of data in question. For example, data related to the (true) effectiveness of a vaccine could well have strong positive externalities, to the extent that it might allow a public health challenge to be addressed. On the other hand, Acemoglu and co-authors argue that there is a negative externality, running from those who disclose data readily and place a low value on privacy to others who value privacy much more highly, but who are similar enough to those who disclose to have their true characteristics possible to infer from the data that is disclosed by others. Hence, the externalities generated by data are mixed, meaning that certain types of data might be overproduced, while other types are underproduced. While we do not focus on personal data issues here, non-personal data are also highly heterogeneous, so the point is

15 See Statista, Statista (2021b). Accessed 1 July 2021.

16 A zettabyte equals 1,000,000,000,000,000,000,000 or one sextillion bytes.

still valid: we cannot conclude from the robust growth of the sector that there is no work to be done.

If we consider informational market failures, increased data access may make markets work better or worse as markets are also highly heterogeneous. Insurance markets are the classic undergraduate example of market failure due to asymmetric information, in either personal or non-personal spheres (depending, in part, on the type of insurance). The scale of this problem is impressive: in 2019, fraudulent claims and dishonest applications amounted to 867,000 cases in the United Kingdom, with an annual value equalling approximately £1.2Bbn[17] in direct costs, but of course preventative measures add to this total. On the other hand, a market currently suffering incomplete but not necessarily asymmetric information could be harmed by adding more information to just one side of the market. While financial markets can benefit from more information to all parties, information to just some of those parties can result in failures, for example.[18]

These remarks assume, of course, that the data that are added to the market are accurate. If the added data are incorrect, then anticipated gains in the functioning of a market may not materialise, and positive externalities may turn into negative externalities. Thinking about vaccines once again, if the information added to the market is incorrect, then decisions may be affected negatively from a social point of view. The quality of the data added to the market matters, whether there is a social oversupply or a social undersupply. Furthermore, data that are accurate but seriously incomplete may not move the market towards a more desirable outcome: we know a great deal about markets with complete information and how they compare to markets with incomplete information, but comparing two markets with asymmetric information, one of which may also be based on false or highly selective information, is less well understood. The nature of the error in the data would affect the conclusion one would reach[19] and whether correcting the error would move the system towards or away from a social welfare improvement.

Finally, even if there are gains to a market due to the increasing use of data, these need not be distributed to consumers. If this is the group whose surplus we aim to maximise, then the benefits of more data across the board are unclear. Consider, for example, data that allow better price discrimination through greater "targeting and customisation." A market with perfect price discrimination may operate very efficiently, but not necessarily to the aggregate benefit of consumers, as all the surplus can be captured by producers.[20] Without accompanying policies, the more efficient market may benefit others than the target group.

Overall, then, if we focus on what any policy change would aim for, the argument for "more data" or "more data access" across the board is less compelling than the argument for more or more access to certain types of data. Further, there is a better argument for more data access to improve the overall efficiency of markets than to improve consumer welfare on its own.

Despite this, for the rest of the article, we limit our attention to socially desirable data only, and how IPRs could be used to improve either the generation of data or data access.

4 Staking Out a Claim

The gold rush involved staking out claims to the source of the gold – initially just a hole in the ground but later plots of land – and obtaining property rights on the gold found on that land. There were two levels of rights in that sense: rights on the source and rights on the fruits of that source. Similarly, one could think of placing rights on the data sources – the people, the sensors, or the experiments that generate the data – or on the data generated by the source (s), or the recorded output. The discussions are linked, but our concern here will be a right for the data itself, as is contemplated in the *sui generis* right for databases.[21]

17 See Statista (2021a) compared to a total contribution to the UK economy of £29.1Bbn, although this was in 2017, reported by a ABI (2019), https://www.abi.org.uk/globalassets/files/publications/public/key-facts/key_facts_2019_spread.pdf. US figures from the Federal Bureau of Investigation (FBI) indicate an estimate of $40Bbn per year in non-health insurance fraud. See Federal Bureau of Investigation (undated) although adding in other types of fraud could well increase this amount considerably: https://www.inguard.com/newsroom/how-insurance-fraud-is-costing-americans-80-billion-a-year/.

18 Goshen and Parchomovsky (2001) argue for a "negative property right" for insider information, which could be thought of in relation to data.

19 For a different application of a similar idea, see Frankel and Rockett's (1988) discussion of international policy coordination when policy-makers not only do not hold the same model but may well be wrong.

20 Personalised markets may also modify bargaining power, possibly to the advantage of customers, but this is not investigated here.

21 The control of those generating the data rather than the data itself has been discussed in the context of the General Data Protection Regulation EU 2016/679, http://data.europa.eu/eli/reg/2016/679/oj.

In this section, we will step back to consider why we have IPRs in the first place and their basic features. We will then see how data do or do not fit into this type of mould. We will explore what this implies for IPR design for data in the next section, after these preliminaries are complete.

Innovation has been cited as a key to economic growth[22] so that it is unsurprising that promoting innovation is enshrined in key documents that establish the European Union, the United States, and others.[23] This is despite the fact that economic theory is rather ambiguous about whether the market underprovides or overprovides innovation compared to the social optimum. We might think that there is too little innovation since innovation can generate a positive externality. For example, when innovations create entirely new markets, inventors neither capture all the surplus generated (in standard posted pricing models), leaving some as consumer surplus, nor do they consistently capture all the benefits of knock-on innovation conducted by rivals. On the other hand, where innovators are private firms or individuals in the same market, they can exert a negative externality on each other when they develop their innovations to "steal business" from each other. For example, a firm may innovate for purely defensive reasons, keeping the innovation away from a rival. Resolving which externality is larger becomes an empirical matter. The work by Bloom et al. (2013) has shown that the empirical evidence[24] points towards social underprovision of innovation, but this is a general empirical regularity, not necessarily true across all markets and time. It is, however, a result that reflects actual behaviour and so will normally reflect all uses, offensive and defensive, of the innovations. In this sense, it is a comforting endorsement of the view that there is too little innovation on average.

Intellectual property protection design is aimed at providing such an incentive to increase innovation. Most IPRs are designed to satisfy a "social contract" that balances the privilege of exclusivity, which is the tool that creates an incentive to innovate, with the obligation to diffuse or make public the protected innovation and the need to avoid excessive deadweight loss in the course of the innovation's use. Standard features, such as term and scope limits on certain IPRs are the tools that balance the interests of innovators, follow on innovators, and the consuming public.

Registration systems are also such a tool. These have many purposes but are also used to facilitate innovation and diffusion as the *quid pro quo* of rights where the nature of the innovation is not revealed adequately by commercialisation. These registration systems, such as the patent register, facilitate follow-on innovation by providing information on the "guts" of an innovation. This will not interfere with the incentive to patent in cases where this registration provides sufficient detail for the basic mechanism of the innovation to be discerned but, at the same time, does not reveal so much that imitation around the right is easy. If "informative registration" does provide the tools to follow on but not imitate, it can direct innovative effort away from perhaps socially wasteful duplication and towards either completely new areas or design-arounds that push the field forward in a socially desirable way.[25]

A second advantage of a registration system is that it publicises the existence of innovation so that it facilitates market trade in innovation via licensing or other tools. These trades allow for wider use, adaptation, and improvement and, following a Coasian logic, can improve social surplus in the process.[26] The registration system aids the creation of a technology market by allowing potential licensing partners to become aware of each other and understand where there can be mutually beneficial trades. While knowledge of potential trading partners often is assumed in modelling analysis of sequential and complementary innovation, this is a feature that must be created and does not occur naturally in real sectors, particularly where the use of an innovation is not a priori known by the innovator (as in the case of undirected innovation), and where partners may be very widely spread geographically and in discipline where existing network connections are thin.

Here, a main concern is privacy regarding personal data. See Acquisti et al. (2016) for a useful overview and references.

22 See, for a useful overview of classic work from a variety of traditions, Verspagen (2006) and the references therein.

23 The Constitution of the United States also enshrines this (Section 8, see The Constitution of the United States), as does the Treaty of Rome in Europe (updated 2002, Title XVI, Article 157(1), OJ 325/103), European Union (1957).

24 That paper draws from US data.

25 See Katznelson and Howells (2021) for a discussion of the difference between "design around" and pure imitation. See Scotchmer and Green (1990), Scotchmer (1991), and references therein for seminal work on sequential innovation. See also Gallini and Wright (1990) or Rockett (1990) for a discussion of information contained in the patent disclosure and its effects on innovation. The type of knowledge that the patent disclosure diffuses can be quite wide, extending beyond the innovation itself (Lee & Lee, 2017).

26 Technology Transfer Guidelines in the European Union and the United States summarise the large literature on licensing by noting that in general it has positive welfare consequences, although with notable exceptions. For the European Union, see European Commission (2014a,b), and for the United States, see Department of Justice and Federal Trade Commission (2017).

Other policies clearly affect this balance of rights between innovator and society. For example, competition policy and other regulation can affect the balance of returns expected from an intellectual property portfolio. Hence, intellectual property policy operates within a web of law, and achieving its aims requires adjusting to changes in other parts of this web to the extent that the expected benefit stream from innovation changes as non-IPR changes occur (such as strengthening of competition law enforcement).[27] Equally, the effectiveness of changes to the design of any one type of intellectual property protection depends on the alternative means of protection: a change in design may just generate substitution away from the instrument. For example, if technical protection measures, such as digital watermarking, or trade secrecy[28] can provide strong protection, then they will be used in preference to other rights where those other rights are difficult to use or do not add further benefit. Hence, the innovative incentives of society depend on a system of rights, not on any single one, and may also depend on access to strategic or technical measures that go well beyond traditional IPRs.

Overall, then, access to information on existing innovations, such as a registry, is an integral part and a *quid pro quo* of certain IPR when it facilitates follow on innovation and matching technology to its various uses. The effectiveness of any one IPR in stimulating innovation depends, however, on how it functions within a web of other laws. These laws combine to affect innovation incentives. Equally, different types of legal, strategic, or technical protection can provide alternatives. Where a right is not easy to use and clearly beneficial, it will not be used, so it will not be an effective tool of intervention. Simply compelling use is not practicable where items such as data are concerned since they are diffuse and may be difficult to monitor. Hence, evaluating any single IPR must be done in conjunction with evaluating the other options facing an innovator and remaining within realistic boundaries of what can be achieved.

5 Gold is not Intellectual Property. Are Data?

We have established from the previous literature that any change we suggest to the current rights framework for data should be straightforward enough to understand and access readily, that certain types of data (those requiring significant investment) should be incentivised and shared (particularly accurate data), that a registry can be a powerful diffusion tool, and that we need to be alive to the incentive to take up the right in the first place. In this section, we consider more closely whether the literature's framework really fits data well enough follow this path.

Using our comparison with gold to focus thoughts, while gold is property, it is not intellectual property. Intellectual property usually only applies to items requiring human effort and creativity, not to "found" items like a nugget lying at the bottom of a river. Moving on to data, a single data point may not reflect much human creativity or even investment, but instead may simply be uncovered, like a piece of gold under a rock, in the course of undertaking normal business activity. At the same time, the European Union has created an IPR-like *sui generis* right to protect databases based on the idea that once data are assembled into a database, it is no longer simply "found" but instead is curated by processing and requiring significant investment to create. This ingenuity and monetary investment would need to be compensated to make production worthwhile. Hence, certain types of data might be considered intellectual property, even if raw data in all their forms are not.

Even if we consider only this subset of all data, however, we still need an idea of how it fits into a framework for modelling innovation. Data are most easily thought of as an input to socially valuable innovation, without intrinsic value in themselves. In this sense, we could consider them a research tool, and perhaps a necessary one: without data, even the parameters of a research plan may not be well defined. While data cannot, then, play the full role of an "innovation" in economic growth models in the sense of generating social surplus on their own, they can trigger an innovation process that does. The benefits accrue from data's offspring, which would not exist without them. The balance of intrinsic benefit to offspring value may be larger for gold, but it, too, is largely an input in other value-generating processes rather than an end to itself.[29]

27 This argument was developed by Regibeau and Rockett (2007). See also Gilbert (2020).

28 See Sandeen and Aplin (2021) for a qualification of what can be understood as trade secrecy. In their exposition, "factual secrecy" captures the idea of effective Trade Secrecy independent of any necessary legal right. Here, we consider trade secrecy as outlined in the European trade secrecy Directive 2016/943, available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32016L0943.

29 Again, note the possibility of holding data in a variety of forms, and potentially benefitting from raw data as a research tool. As we are focussed here on IPR modifications, we do not consider raw data in any detail, but we do note below the possibility of substituting into it

This role of data in a sequential stream of innovation is a form of complementarity: as a research tool, data are complementary to their own offspring. Data also may be complementary to each other. For example, dataset A may not be necessary to collect dataset B; however, using datasets A and B together can produce insights that were not possible with either dataset on its own. To make things more concrete, it may not be possible either to evaluate the quality of existing disease treatments or to improve on those treatments without data on patient performance after exposure to a particular therapy. These complementarities mean that data can generate positive externalities (without access to data, further innovation may not be possible) or negative externalities (datasets that must be used in combination to generate innovation create a strategic incentive for each dataset to be priced too high[30]).

Hence, data can be an innovation market gatekeeper[31] or at least a facilitator in either a sequential innovation framework or in a framework where they form a necessary element in a larger data-gathering exercise if they are closely held or non-replicable from other data sources. Indeed, the source of such data may be a smaller or a larger firm, so it may be quite difficult to situate such a gatekeeper or a facilitator research tool in an industrial landscape. In the more general IPR literature, broader legal protection of research tools, allowing them to capture some of the profit from follow-on innovations, has been proposed to solve the problem of incentivising socially optimal creation of research tools in the presence of their positive externality on further innovation.[32] This solution does not address the case where multiple datasets are necessary to create follow-on innovation. To resolve this problem, the literature suggests pooling the necessary elements so that they can be priced together into a single accessible package, usually licensed under FRAND commitments,[33] although the work that has been done so far is not tailored to the rather unique case of data pools.

Where broad protection is used to allow data holders to capture the positive externality on future innovation that they generate, the sequential innovation literature indicates that the research tool will be shared by relying on a Coasian logic as long as property rights are clear and the market for technology (and data) functions efficiently enough.[34] While this seems straightforward, there are several areas where data throw up challenges. Firstly, it is not clear that trade is what we observe: inadequate data diffusion tends to be a more general complaint than inadequate data generation. One possibility is that data are not currently protected strongly enough to create a rights regime that allows the Coase Theorem to work. Another is that data are protected strongly enough, but other assumptions of the sequential innovation literature that generate conclusions about the optimality of strong rights do not hold in the case of data. Here, the sequential innovation literature generally focuses on the case where the data holder must pair with other independent entities to generate an innovation (Scotchmer, 1991, 1996, for example). This "scarcity of ideas"[35] and attendant need to pair with others for innovation to occur may be true in some cases, but it is not uncommon for a data holder to be an both a data generator and an innovator in its own right, able to use either factual or legal (or both) trade secrecy to protect the data and delay the entry of others. Indeed, the expertise to collect the data often is associated with the ability to use it. Hence, it is not clear that the case of "idea scarcity" applies to the datasphere. If it does not, then the incentive to trade is not as strong because concealing the innovation entirely offers an attractive alternative. Even if it does, it is not entirely clear how to carry over the idea of large leading breadth within a technology trajectory to the case of data. One possible solution is to use grantbacks to achieve similarly broad protection for data.

as a way of avoiding any unpalatable regulations. The emphasis really needs to be on both creating an attractive right and managing the alternative.

30 This has been discussed extensively under the rubric of "royalty stacking," see Lemley and Shapiro (2007). The problem of "Cournot Complements" more generally can lead to overpricing of inputs. See Spulber (2017) for more general discussion.

31 This is not the same as the definition of a gatekeeper in the Digital Markets Act (see Regulation (EU) 2022/1925, Article 3), European Parliament and Council (2022b) as we are concerned with innovation markets, not product markets. In this sense, it is closer to the role of data as an innovation technology as in the study by Prufer and Schottmuller (2021). Here, however, we focus on a subset of data, that requiring investment to create, which is one point of distinction between our work and theirs.

32 For this to work smoothly, *ex ante* licensing must be both possible and legal. See Green and Scotchmer (1995) and also see Ambashi et al. (2019) for a discussion of grantbacks to address this issue and Choi (2002) for further discussion.

33 The appropriate design of the patent pool was discussed by Lerner and Tirole (2004), where it is also pointed out that whether individual elements of the pool are complements or substitutes may depend upon the price of the element. This ambiguity will also be present in data. At the same time, that article was not adapted to the specific case of data, which might change the design recommendations. For example, data pooling may not be around a single design, as pooling for the purposes of standard setting would be.

34 For example, see Green and Scotchmer (1995).

35 Scotchmer (2004) expands on the idea of ideas being scarce.

A second feature of data, other than the potential for the data holder to also be innovator, is that data often are held as a trade secret, as we saw in Figure 2. Datasets need not be revealed in their use, in contrast to a book or a film, which must be revealed if it is to create value at all.[36,37] The fact that secrecy – without trade – is a viable and apparently attractive alternative for protecting data has two implications where data are also a research tool in a sequential innovation process. Firstly, if data are relatively easy to hide whilst being used to create value, then uptake of diffusion is by no means guaranteed. This suggests that the incentive to take up a right that is associated with the revelation of the existence of the data may be far less obvious than it is for product innovations that would be revealed in the normal process of their exploitation.

This substitution into trade secrecy – if it is not associated with licensing and instead can be thought of as concealment – may come at a social cost. While protecting an innovation using trade secrecy allows for a potentially desirable independent invention system, as advocated by Maurer and Scotchmer (2002) and emphasised by Cugno and Ottoz (2006), and can be optimal in an *ex post* sense once innovation exists it does not have the positive effects of directing innovation, as does the patent system. This direction comes through the patent disclosure, which can be used to identify business opportunities (Lee & Lee, 2017), learn about competition and so optimally adjust research plans, or use the patent as a springboard to socially beneficial "invent arounds" (Katznelson & Howells, 2021). In other words, the diffusion aspect of the patent register is important to innovation's trajectory and is not part of trade secrecy and certainly not part of a strategy that relies on concealment.[38] Rather, secrecy may incentivise duplicative research expenditure, even though the social value of duplication is debatable: socially, it could be a far better use of resources to diffuse an innovation than to spend to reinvent it. Further, concealment combined with the ability of the data holder to also generate innovations flowing from the data can mean that data are released too slowly and innovations follow with more delay than the social optimum.[39] The lack of disclosure creates, then, a potential loss for a society where coordinating efforts across firms is important to generate innovation. Further, without a term limit, trade secrets can allow for a long period of both exclusive use of an innovation and lack of disclosure. Hence, the lack of a term limit in the case of secret material has two, not one, source of loss.

Hence, data can serve as an important research tool with both forward complementarities to innovations and "sideways" complementarities to other datasets, both of which create externalities that have been pointed out in earlier literature on IPRs as requiring intervention either in rights design or in access pricing and pooling design. Optimal rights can be quite strong, as early innovations exert a positive externality on future innovations. Still, the IPR literature's conclusion about strong protection relies upon assumptions that partners are easy to identify, ideas are scarce, and incentives for trading exist. These assumptions may not hold for data. Further, the literature often has analysed the IPR design of single instruments in isolation from the portfolio of options that actually exist. To address the case of encouraging data creation and diffusion, we may need a system that creates the knowledge of the existence of partners in a credible way and also provides an incentive for trading in the light of quite attractive non-trading options.

6 If Data are Data, not Gold, What Does It Mean for Property Rights?

One can make several comments about the features of a protection system for data based on these observations. These distinctive features can form the basis for adjusting the rights regime.

Firstly, diffusion via a market cannot develop if agents do not know with whom they should partner.[40] While the Data Governance Act provides a registration system for data traders and suggests that interoperability be improved, it

36 Perhaps analogously, Cohen et al. (2000) report that process innovations tend to be protected more frequently with trade secrecy than product innovations.

37 Data may be easier to steal as well, and if this is done in a decentralised manner, it may be hard to detect. For this reason, technical protection measures may also be favoured for data since they make data less transportable. While we focus mainly on legal protection mechanisms here, this is by no means the only alternative for data.

38 See Paine (1991).

39 See Matutes et al. (1996) for background on this argument. That article shows that an initial innovator will have an incentive to delay publication until some further "follow-on" innovations are developed in house. Exploiting trade secrecy allows the data holder to capture a larger share of the entire stream of innovations that may result from the data but can delay the innovation process. This delay represents a social loss.

40 Using a market to govern diffusion allows data that are more useful for further innovation to command a higher price, so that the market can signal where it is more useful to put one's effort into collecting data in the first place.

does not provide any kind of catalogue, or index, of the contents that are available for trade. This is the more relevant information to reduce the barriers to the formation of a vibrant market.

A registration system that creates a central repository for data holdings, eventually on a worldwide basis, could promote socially beneficial trading to develop across entities that are not known to each other *ex ante*. Trade secrecy – or any other type of protection – as long as registration is added to the protection – does not negate this benefit, although it may reduce uptake, as we discuss in the following sections. Making such a system as comprehensive as possible allows for the best opportunities for matching up complementary data, particularly as new artificial intelligence innovations require large amounts of varied training data to work well. There is no requirement to register databases that are protected with the Database Directive at present.[41]

While various worldwide repositories exist for limited types of data, there is no currently agreed protocol for worldwide holdings, access, or a single place to find all data. The market incentives to develop a single system exist but may not generate such harmonisation quickly.[42] Market-based data sharing platforms exist and have developed rapidly,[43] but it is not clear that the market and social incentives for interoperability and sites that allow comparison shopping for similar datasets exists. Recent works by Delbono et al. (2021) and Martens and Mueller-Langer (2020) suggest that the data broker market may be subject to imperfections and problematic (exclusionary) behaviour. Both papers emphasise that analysis of the data broker market is in its infancy; however, it is unclear whether regulation of the existing market or government sponsorship dominates. The decision of whether to regulate what is already there or undertake full government sponsorship depends on factors, such as the speed of the regulatory or government-sponsored implementations and the severity of the market failure compared to any government failure. Full government sponsorship is one way, but not necessarily the only way, to create such a unified, interoperable, system and, indeed, the EDIB has interoperability standards as one item in its remit. Whatever system was developed, it would need to be straightforward to use and access to be useful, of course.

Creating a repository that promotes trade requires that entries reveal the nature and existence of the dataset such that contracting for trade can proceed without revealing the full contents and so destroy the incentive to participate. This difficulty was pointed out by Arrow as an information paradox: contracting on information is impossible if it requires disclosure to conclude a contract, but this revelation renders a contract to reveal it unnecessary. Kerber (2016) argues that meta-data or other summary information makes this eminently possible. Indeed, cataloguing systems that describe, classify, and cross reference the data down to relatively specific subject matter are already used by data libraries. Further, such classifications provide much of the ancillary information on technological categories that can facilitate search and comparison and creation of systems of complementary data. Note, too, that the concerns of Anton and Yao (2004) on bias in the selection of material that is deposited to a patent repository do not necessarily hold for such meta-data, as this description does not enable replication of the data, but merely enables trade so as to enable innovation built upon the data. In this sense, a "market for lemons" problem that might be expected among patents may not exist for a data repository. Of course, those who wish to make the full dataset open access can do so, and a comprehensive repository also helps them to find users.[44]

An additional feature that could improve the fluidity of the market for data is quality certification of registered data. The accuracy of a dataset and true identity of the owner would promote diffusion of data that are likely to contribute to rather than detract from social welfare. Curation according to standard protocols is already recommended in the European Union for certain types of data,

41 While some of the acts in Figure 1 covering data do require a certain amount of access, the point of this article is that the restrictions on the categories (such as public sector data only or restricting attention to firms that are market gatekeepers) excludes categories of data that are and will likely become increasingly important to access in the years to come. We aim here at a unified system for all processed data, not just that of certain sectors or entities.

42 The Data Governance Act in the European Union (see European Parliament and Council (2022a)) proposes that interoperability be encouraged and that work continue on protocols. Currently proposed legislation in India on a repository for non-personal data exists, but that proposal involves other aspects that would not be supported by this argument, such as open access. For a summary of the proposals, see: https://www.snrlaw.in/personal-and-non-personal-data-in-digital-india-before-and-after/.

43 While Kerber (2016) notes that development is not yet vigorous, more recent surveys, such as https://www.maximizemarketresearch.com/market-report/global-data-broker-market/55670/, suggest a market growing at a projected 4.5% per annum, albeit mainly trading in personal data and including both processed and unprocessed (raw) data.

44 This goes beyond the recommendation to facilitate trade by publicising data exchange agents. Such agents can be complementary to such a system, but the repository allows for direct contact, avoiding these middlemen if one so desires.

and is addressed – albeit somewhat obliquely – in Article 12 (paragraph g) of that Act. The recommendation here is to firm up this element to create a minimum government-sponsored quality standard that is linked to registration. Full government sponsorship would improve the chance of trustworthy certification, although industry standards could be substituted. It is not clear, of course, that market incentives and social incentives for quality are the same. The expertise developed to provide certification services could improve government's ability to act on any market-based initiatives, including uncovering fraudulent data or unreliable data-sharing platforms. Unethical practices such as "false advertising" of poor quality data as high quality or simply trading in low-quality data are condemned in the Data Governance Act, but enforcement requires some form of "consumer protection" bureau for data. Enforcement requires both information on the existence of data and expertise to evaluate the data. Enforcement certainly could be facilitated by the requirement to submit, at the time of registration, sufficient information for the certifying agency to verify that the data are what they say they are, that they are legitimately the property of the purported owner, and that the owner is who is claimed would facilitate this. A single clearinghouse with indexed information that also gives ownership could also allow data holders to check for piracy.

The benefits of participating in a large network may not be enough to incentivise data holders to participate, however. While entry can be a *quid pro quo* of protection under the *sui generis* right, not all users will wish to protect in this way. The popularity of trade secrecy – without any requirement to reveal the existence or nature of that data – suggests that further hooks would be needed.

The social loss from concealment compared to registration derives from the absence of information about partners, on which further innovation could build, along with any ancillary benefits in terms of detecting piracy. A second loss is that there is no expiration of the protection term as long as the data are kept secret. Requiring registration of any data receiving protection would require a data exception to either trade secret coverage, as registration is not required and full concealment is possible whilst retaining full rights under trade secrecy. In other words, unregistered data would remain accessible either through direct sharing (because the information loses the benefit of confidentiality) or a freedom of information request (if the material is not directly shared). This type of exception would generate full protection to any data that are declared through the registry, including possibly access to grantbacks, but very little protection to that which is not declared. A similar argument could be made with respect to technological protection measures or any other measure that is undertaken whilst maintaining the data secret. Article 30, paragraph (d) of the Data Governance Act allows for trade secrecy and other intellectual property protection to be reviewed by the EDIB in the case of data, and the data exception suggested here would be one avenue to take this remit.[45]

This is not quite the same as a requirement that data be "worked" to maintain any right to exclusivity. The requirement to work does not achieve revelation of the data's existence and so does not promote the development of a marketplace. There is a difference between gold and data in this respect: revelation of the existence of a land claim was "automatic" in the case of gold because the land could be physically observed by anyone who wished to travel to the area. In the case of data, the claim itself can remain hidden, regardless of whether it is worked. The fact that the working requirement was added successfully to gold claims does not, then, necessarily mean that it would be helpful for the case of data. At the same time, given that data do not necessarily fulfil all the assumptions in the sequential innovation literature on scarcity of ideas, some attention might need to be given to whether data should lose its protection where it is not worked after a certain period of time. Again, this is a potential exception to trade secrecy and could also fall within the remit of the EDIB.

This gives the main strokes of the components of a modified data right: registration in a way that allows the largest network to gather and creates an accessible index of contents, a sponsored standard of quality certification that allows users to be confident of what they are buying and from whom, reduced protection of data where they are not registered, and a potential exception for working the data are all modifications that could be considered.

These are not the only elements that deserve attention, but they are perhaps the most pressing. FRAND provisions are a standard tool to control pricing of complementary inputs to innovation, although this system is not without its critics, and is already present. Ensuring that this applies to pools of complementary datasets, however, would be an additional step to make explicit.

Some types of data may be held in many forms, including raw data and processed data forms of the same material. While it is not a focus of this piece, one would not want to design obligations that apply to one way of holding

45 This does not mean that "every little thing" would end up being registered. Protection is lower for unregistered data, but equally most data are not very useful to others, and so there is no threat for these data if the lower protection is selected.

data and simply create substitution into another. Raw data, as it does not fall under existing intellectual property rules, would likely need to be concealed to be able to extract any surplus. Further, if the raw data are obtained in the course of doing business and do not require any significant investment would not require the same design of protection as data requiring investment: there is no need to incentivise creation in this case, as it occurs automatically. For these types of data, minimal protection is necessary, as the main concern is to diffuse what is available as easily as possible. This is visible in the provisions of the Data Act.

All these suggestions for rights design are not static: they would need to be modified as data uses develop. A case in point is artificial intelligence. This technology can place more value on highly granular data, as such data can aid in the learning and inference process. This would require matching granularity in the data registration system. Further, to the extent that artificial intelligence systems may shift the balance of research towards less directed and more undirected research due to the capacity of such systems to generate unexpected solutions, a single dataset may generate more far-flung follow-on innovations than might have been anticipated. This suggests greater importance for knowledge of and potential access to data, as well as mechanisms to facilitate their trade (such as grantback and FRAND provisions) in the future, shifting the balance between the benefits of access and those of exclusivity for innovation. Of course, artificial intelligence also poses other issues for policy, not the least of which is whether a human "minder" should always be associated with such programmes to ensure that substantive penalties could be enforced on innovations that either generate abuse or outright break the law.

A general comment on all these directions for protection would be that any protection system for data would be most useful if undertaken on a worldwide basis and for a wide variety of data. Data users can be widely dispersed. At the same time, a worldwide system immediately runs afoul of no standard set of understandings on what data (and datasets) are, and no standard law surrounding disclosure of the information (such as privacy rights). This work would need to be undertaken in conjunction with work on designing a new property right with a worldwide forum, such as WIPO, as the most natural sponsor.

7 Conclusions

Data are data, and gold is gold. They are not the same and should not be protected in the same way. At the same time, the protection of gold claims in the Gold Rush of the mid-nineteenth century in the United States faced a similar problem of creating rights to a newly discovered resource amidst a flurry of activity. During the Gold Rush, existing legislation was adapted, sometimes with several changes before an acceptable model was revealed, to address this new problem. In the case of data, the same approach is possible. Current rights are available, such as the *sui generis* right and the framework of the Data Governance Act in the European Union but may need to be adapted further to obtain better results for society.

We have suggested some adjustments based on the literature on innovation and intellectual property design to accommodate to the developing datasphere and data's evolving uses and methods of use including in the area of artificial intelligence. Registration and cataloguing with a uniform method of holding and accessing entries, quality monitoring and minimum certification by a trusted source, and modifications of alternative rights into which data holders might substitute all are potential changes that could improve diffusion whilst preserving the incentive to invest. Important issues such as which datasets would be susceptible to any such changes need to be considered, of course. This article is focussed on the broad strokes of modifications that could be considered, which could feed into thinking on data governance, perhaps finding their way into an expanded remit for the EDIB. This would allow the knotty issues around them, including enforcement, eligibility, and implementation, to be considered in full. Further, many of the issues have been modelled only partially in the economics literature and so deserve more thorough investigation.

Acknowledgments: A preliminary version of this article was presented as a keynote talk to the MPI Alumni Conference, July 9, 2021. Many thanks to the audience at that conference, Suma Athreye, Nicola Searle, and two anonymous referees for their helpful comments.

Funding information: The author states no funding involved.

Conflict of interest: The author is editor-in-chief of the journal but was not involved in peer-review and decision making regarding the article.

Article note: As part of the open assessment, reviews and the original submission are available as supplementary files on our website.

References

Acemoglu, D., Makhdoumi, A., Malekian, A., & Ozdaglar, A. (2022). Too much data: Prices and inefficiencies in data markets. *American*

Economic Journal: Microeconomics, *14*(4), 218–256. doi: 10.1257/mic.20200200.

Acquisti, A., Taylor, C., & Wagman, L. (2016). The economics of privacy. *Journal of Economic Literature*, *54*(2), 442–492. doi: 10.1257/jel.54.2.442.

Ambashi, M., Regibeau, P., & Rockett, K. (2019). Grantbacks, territorial restraints, and innovation. *International Journal of Industrial Organization*, *67*, 102534. doi: 10.1016/j.ijindorg.2019.102534.

Anton, J., & Yao, D. (1994). Expropriation of inventions: Appropriable rents in the absence of property rights. *American Economic Review*, *84*(1), 190–209.

Anton, J., & Yao, D. (2004). Little patents and big secrets: Managing intellectual property. *Rand Journal of Economics*, *35*(1), 1–22.

Association of British Insurers (ABI). (2019). *UK insurance and long term savings key facts [Fact Sheet]*. Retrieved November 1, 2021, from https://www.abi.org.uk/globalassets/files/publications/public/key-facts/key_facts_2019_spread.pdf.

Bessen, J. (2005). Patents and the diffusion of technical information. *Economics Letters*, *86*(1), 121–128. doi: 10.1016/j.econlet.2004.07.005.

Bessen, J., & Maskin, E. (2009). Sequential innovation, patents, and imitation. *Rand Journal of Economics*, *40*(4), 611–635. doi: 10.1111/j.1756-2171.2009.00081.x.

Bloom, N., Schankerman, M., & Van Reenen, J. (2013). Identifying technology spillovers and product market rivarly. *Econometrica*, *81*(4), 1347–1393. doi: 10.3982/ECTA9466.

Boyle, J. (2006, January 2). *Two database cheers for the EU*. Financial Times.

Choi, J. P. (2002). A Dynamic Analysis of Licensing: The 'Boomerang' Effect and Grant-back Clauses. *International Economic Review*, *43*(3), 803–829. doi: 10.1111/1468-2354.t01-1-00036.

Clay, K., & Wright, G. (2005). Order without Law? Property Rights during the California Gold Rush. *Explorations in Economic History*, *42*(2), 155–183. doi: 10.1016/j.eeh.2004.05.003.

Cohen, W., Nelson, R., & Walsh, J. (2000). Protecting Their Intellectual Assets: Appropriability Conditions and Why US Manufacturing Firms Patent. *NBER Working Paper 7552*.

Cugno, F., & Ottoz, E. (2006). Trade secret vs broad patent: The role of licensing. *Review of Law and Economics*, *2*(2), 209–221. doi: 10.2202/1555-5879.1069.

Delbono, F., Reggiani, C. & Sandrini, L (2021). Strategic data sales to competing firms. *JRC Digital Economy Working Paper 2021-05, European Union.* https://joint-research-centre.ec.europa.eu/system/files/2021-12/jrc126568.pdf.

Deloitte. (2018). *Data is the new gold: The future of real estate service providers*. Retrieved November 1, 2021, from https://www2.deloitte.com/global/en/pages/real-estate/articles/future-real-estate-data-new-gold.html.

Dosi, G., Marengo, L., & Pasquali, C. (2006). How much should society fuel the greed of innovators?: On the relations between appropriability, opportunities and rates of innovation. *Research Policy*, *35*, 1110–1121. doi: 10.1016/j.respol.2006.09.003.

Drexl, J., Hilty, R., Desaunettes, L., Greiner, F., Kim, D., Richter, H., Surblytė, G., & Wiedemann, K. (2016). Data ownership and access to data: Position statement of the Max Planck Institute for Innovation and Competition of 16 August, 2016 on the Current European Debate. *Max Planck Institute for Innovation and Competition Position Paper*. Retrieved November 1, 2021 from www.ip.mpg.de/fileadmin/ipmpg/content/stellungnahmen/positionspaper-data-eng-2016_08_16-def.pdf.

European Commission. (2014a). *Regulation 316/2014 21 March 2014 on the Application of Article 101(3) of the Treaty of the Functioning of the European Union to Categories of Technology Transfer Agreements (TTBER)*. OJ L93, 28.03.2014. 17–23.

European Commission. (2014b). *Guidelines on the Application of Article 101 of the Treaty on the Functioning of the European Union to Technology Transfer Agreements (Guidelines)*. OJ C89, 28.03.2014. 3–50.

European Commission. (2018). *Study in Support of the Evaluation of Directive 96/9/EC on the Legal Protection of Databases Annex 2: Economic Analysis. SMART number 2017/0084*. Retrieved November 1, 2021 from https://op.europa.eu/en/publication-detail/-/publication/5e9c7a51-597c-11e8-ab41-01aa75ed71a1/language-en.

European Commission. (2020). *The European data market monitoring tool: Key facts and figures, first policy conclusions, data landscape, and quantified stories. D2.9 Final Report*. Retrieved November 1, 2021 from https://op.europa.eu/en/publication-detail/-/publication/9fb0599f-c18f-11ea-b3a4-01aa75ed71a1/language-en.

European Parliament and the Council. (1996). *European Parliament and Council Directive (EC) 96/9/EC On the Legal Protection of Databases*. OJ L77/20.

European Parliament and the Council. (2022a, 30 May). *Regulation (EU) 2022/868 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act)*. OJ L152/1. 3.6.2022.

European Parliament and the Council. (2022b, 14 September). *Regulation (EU) 2022/1925 on contestable and fair markets in the digital sector and amending Directives (EU) 2019/1937 and (EU) 2020/1828 (Digital Markets Act)*. OJ L265/1. 12.10.2022.

European Union. (1957, 25 March). *Treaty Establishing the European Community (Consolidated Version), Rome Treaty*. Retrieved November 1, 2021 from https://www.refworld.org/docid/3ae6b39c0.html and updated 2002 as consolidated version of the treaty establishing the European community. OJ. 325/35, 24/12/2002.

Federal Bureau of Investigation. (undated). *Insurance Fraud. Federal Bureau of Investigation Reports and Papers*. Retrieved November 1, 2021 from https://www.fbi.gov/stats-services/publications/insurance-fraud.

Fontana, R., Nuvolari, A., Shimizu, H., & Vezzulli, A. (2013). Reassessing patent propensity: Evidence from a dataset of R&D Awards, 1977–2004. *Research Policy*, *42*, 1780–1792. doi: 10.1016/j.respol.2012.05.014.

Forbes Africa. (2019). Data is the new gold. *Forbes Africa*. Retrieved November 1, 2021 from https://www.forbesafrica.com/technology/2019/07/18/data-is-the-new-gold/

Fosfuri, A., & Ronde, T. (2004). High tech clusters, technology spillovers, and trade secret laws. *International Journal of Industrial Organization*, *22*(1), 45–65. doi: 10.1016/S0167-7187(03)00123-1.

Frankel, J., & Rockett, K. (1988). International macroeconomic policy coordination when policy-makers disagree on the model. *American Economic Review*, *78*(3), 318–340.

Gallini, N., & Wright, B. (1990). Technology transfer under asymmetric information. *Rand Journal of Economics*, *21*(1), 147–160. doi: 10.2307/2555500.

Gans, J., Hsu, D., & Stern, S. (2007). The impact of uncertain intellectual property rights on the market for ideas: Evidence from patent grant delays. *NBER Working Paper 13234*.

Gilbert, R. (2020). *Innovation matters: Competition policy for the high technology economy*. MIT Press.

Goldfarb, A., & Tucker, J. (2019). Digital economics. *Journal of Economic Literature*, *57*(1), 3–43. doi: 10.1257/jel.20171452.

Goshen, Z., & Parchomovsky, G. (2001). On insider trading, markets and 'negative' property rights in information. *Virginia Law Review*, *87*(7), 1229–1277.

Graef, I., & Prufer, J. (2021). Governance and Data Sharing: A Law and Economics Proposal. *Research Policy*, *50*(9), 104330. doi: 10.1016/j.respol.2021.104330.

Green, J., & Scotchmer, S. (1995). On the division of profit in sequential innovation. *Rand Journal of Economics*, *26*(1), 22–33.

Halloran, D., & D'Souza, F. (2020). Data is the new gold: This is how it can benefit everyone – While harming no one. *World Economic Forum*. Retrieved November 1, 2021 from https://www.weforum.org/agenda/2020/07/new-paradigm-business-data-digital-economy-benefits-privacy-digitalization/.

Heger, D., & Hussinger, K. (2017). Implications of uncertain patent rights for German start-ups' commercialisation activities and access to external capital. *Industry and Innovation*, *24*(7), 753–773. doi: 10.1080/13662716.2016.1264066.

Katznelson, R., & Howells, J. (2021). Exclusive rights stimulate design around: How circumventing Edison's lamp patent promoted competition and new technology development. *Journal of Competition Law Economics*, *17*(4), 1007–1052. doi: 10.1093/joclec/nhab010.

Kerber, W. (2016). A new (intellectual) property right for non-personal data? An economic analysis. *Gewerblicher Rechtsschutz und Urheberrecht, Internationaler Teil (GRUR Int)*, *65*(11), 989–999.

Lee, M., & Lee, S. (2017). Identifying new business opportunities from competitor intelligence: An integrated use of patent and trademark databases. *Technological Forecasting and Social Change*, *119*, 170–183. doi: 10.1016/J.TECHFORE.2017.03.026.

Lemley, M., & Shapiro, C. (2007). Patent holdup and royalty stacking. *Texas Law Review*, *85*, 1991–2049.

Lerner, J., & Tirole, J. (2004). Efficient patent pools. *American Economic Review*, *94*(3), 691–711. doi: 10.1257/0002828041464641.

Martens, B., & Mueller-Langer, F. (2020). Access to digital car data and competition in the aftermarket maintenance services. *Journal of Competition Law & Economics*, *16*(1), 116–141. doi: 10.1093/joclec/nhaa005.

Matutes, C., Regibeau, P., & Rockett, K. (1996). Optimal patent design and the diffusion of innovations. *Rand Journal of Economics*, *27*(1), 60–83.

Maurer, S., & Scotchmer, S. (2002). The independent invention defense in intellectual property. *Economica*, *69*(276), 535–547. doi: 10.1111/1468-0335.00299.

Moser, P. (2012). Innovation without patents: Evidence from nineteenth century world's fairs. *Journal of Law and Economics*, *55*, 43–74.

Moser, P. (2013). Patents and innovation: Evidence from economic history. *Journal of Economic Perspectives*, *27*(1), 23–44. doi: 10.1257/jep.27.1.23.

Paine, L. (1991). Trade secrets and the justification of intellectual property: A comment on Hettinger. *Philosophy & Public Affairs*, *20*(3), 247–263.

Prufer, J., & Schottmuller, C. (2021). Competing with big data. *Journal of Industrial Economics*, *LXIX*(4), 967–1008.

Regibeau, P., & Rockett, K. (2007). The relationship between intellectual property law and competition law: An economic approach. In S. Anderman (Ed.), *The interface between intellectual property rights and competition law* (pp. 505–552). Cambridge University Press.

Reichman, J., & Samuelson, P. (1997). Intellectual property rights in data. *Vanderbilt Law Review*, *50*, 49.

Richter, H., & Slowinski, P. R. (2019). The data sharing economy: On the emergence of new intermediaries. *IIC*, *50*, 4–29. doi: 10.1007/s40319-018-00777-7.

Risch, M. (2007). Why do we have trade secrets?. *Marquette Intellectual Property Review*, *11*, 1–76.

Rockett, K. (1990). The quality of licensed technology. *International Journal of Industrial Organization*, *8*(4), 559–574.

Sandeen, S., & Aplin, T. (2021). Trade secrecy, factual secrecy and the hype surrounding AI. In R. Abott (Ed.), *Research handbook on intellectual property and artificial intelligence* (pp. 443–461). Edward Elgar.

Scotchmer, S. (1991). Standing on the shoulders of giants: Cumulative research and patent law. *Journal of Economic Perspectives*, *5*(1), 29–41.

Scotchmer, S. (1996). Protecting early innovators: Should second-generation products be patentable? *The Rand Journal of Economics*, *27*(2), 322–331. doi: 10/2307/2555929.

Scotchmer, S. (2004). *Innovation and incentives*. MIT Press.

Scotchmer, S., & Green, J. (1990). Novelty and disclosure in patent law. *Rand Journal of Economics*, *21*(1), 131–146.

Smith, H. J., Dinev, T., & Xu, H. (2011). Information privacy research: An interdisciplinary review. *MIS Quarterly*, *35*, 989–1015.

Spulber, D. (2017). Complementary monopolies and bargaining. *Journal of Law and Economics*, *60*(1) 29–74.

Statista. (2021a). *Number of Fraudulent Insurance Claims and Dishonest Applications Detected in the United Kingdom (UK) from 2017 to 2019. Statista*. https://www.statista.com/statistics/885271/fraudulent-insurance-claims-and-dishonest-applications-united-kingdom/.

Statista. (2021b). *Annual Size of Real Time Data in the Global Datasphere from 2010 to 2025. Statista*. https://www.statista.com/statistics/949144/worldwide-global-datasphere-real-time-data-annual-size/.

The Constitution of the United States. Retrieved November 1, 2021, from https://constitutioncenter.org/interactive-constitution/full-text.

US Department of Justice and Federal Trade Commission. (2017). *Antitrust guidelines for the licensing of intellectual property*. Retrieved November 1, 2021 from https://www.ftc.gov/system/files/documents/public_statements/1049793/ip_guidelines_2017.pdf.

Verspagen, B. (2006). Innovation and economic growth. In J. Fagerberg & D. Mowery (Eds.), *Oxford handbook of innovation* (pp. 487–514). Oxford University Press.

World Intellectual Property Organisation, WIPO. (2002). *Summary on Existing Legislation Concerning Intellectual Property in Non-Original Databases*. World Intellectual Property Organisation. SCCR/8/3. Retrieved November 1, 2021, from https://www.wipo.int/edocs/mdocs/copyright/en/sccr_8/sccr_8_3.pdf.