Economics

Is Data the New Gold? Considering Intellectual Property Protection of Data*

--Manuscript Draft--

Manuscript Number:	ECONJOURNAL-D-21-00134
Full Title:	Is Data the New Gold? Considering Intellectual Property Protection of Data*
Article Type:	Research Article
Keywords:	data, intellectual property protection, trade secrecy, registration
Manuscript Region of Origin:	UNITED KINGDOM
Abstract:	The paper reviews the economics literature on the optimal design of intellectual property, the current framework for protecting data (largely in the EU) and the main features of data. It combines these to ask how intellectual property protection instruments could be adapted to usefully protect (largely processed) data. The main changes from current protection that are suggested include an expansion of registration and cataloguing systems and increased attention to the attractiveness of trade secrecy as an alternative protection mechanism.
Manuscript Classifications:	8.5: Publicly Provided Goods; 15.3: Innovation • Research and Development • Technological Change • Intellectual Property Rights

Is Data the New Gold? Considering Intellectual Property Protection of Data*

2 December, 2021

Katharine Rockett

Abstract

The paper reviews the economics literature on the optimal design of intellectual property, the current framework for protecting data (largely in the EU) and the main features of data. It combines these to ask how intellectual property protection instruments could be adapted to usefully protect (largely processed) data. The main changes from current protection that are suggested include an expansion of registration and cataloguing systems and increased attention to the attractiveness of trade secrecy as an alternative protection mechanism.

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

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 others[1]. It is an appealing image. The gold rushes of the 19th 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. If one carries through the analogy, one quickly finds property rights were at the heart of the behaviour of agents during the Gold Rush. Clay and Wright (2003) study the records of the US Gold Rush, beginning in 1848, finding that claims to specific property rights were established soon after the initial rush and 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. This 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 and adaptations of earlier laws, notably 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 with improvement conditions attached.

What can we say, then, about property rights to data and how these should be designed to get the most for society out of data? There are many similarities between data and gold in the feeling of a "rush", but significant differences as well that 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 need not be the same. The analogy has quite limited usefulness once one examines the subject matters involved.

The purpose of this paper is to look at data's features, consider the data rights landscape as it currently exists (in Europe in particular) and, starting from a base of standard intellectual property rights (IPRs), suggest adaptations to these that might help to accommodate data's unique features. In particular, processed data is a focus of the paper, with the moderate performance of the *sui generis* database protection as a point of departure[2]. Similarly, the paper skirts the issues of protecting the sources of data or restrictions (mainly privacy concerns) on the curation and use of data or with transparency issues such as profiling. Rather, the paper addresses some dimensions of a useful property right, drawing from but perhaps not entirely like an IPR, for processed data including some areas to prioritise and some areas to postpone. The legislative landscape of rights is both complex and new, and some

[1] See Halloran and D'Souza (2020), Forbes Africa (2019) and Deloitte (2018) as some of many examples available on an internet search.
[2] See Graef and Prufer's (2021) summary of recent EU legislation on raw data sharing and their argument for mandatory data sharing for raw user data. As Graef and Prufer (2021) note, processed data involves an investment and normally should be treated distinctly. Despite the starting point of European systems in both this paper and theirs, the mechanisms envisaged here are not suggested for Europe only.

of these rights may well be simplified and pruned in the coming years. For this reason, a reasonably simple framework for prioritising may be timely.

Other work has been done on this issue: the European Commission (EC) has published a set of consultation papers on the Database Directive that also aims to consider pruning and adjustment (or, indeed, elimination) of the right. (WIPO, 2002) summarises progress in various countries on protection of non-original data that notes the variety of features in national legislation to give a menu for design and adjustment. This paper, in contrast, focuses on the conceptual arguments for intellectual property right design, rather than on the empirical aspects of the landscape. It also makes some distinct points, with the "default" protection and how to incentivise agents to take up any data protection rights emphasised here. Indeed, in some sense this paper is a response to the puzzle in the EC consultation paper on why Database protection is not used more and what might affect its take-up to the benefit of social welfare. In contrast to the work of Graef and Prufer (2021), we aim for a simple system.

This paper will argue most strongly for a registration system for processed (meta)data as an initial step to facilitate trading: 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 one-stop shop. This is not a decentralised system of registration of data traders, as has been considered[3], but rather a centralised system to register the (meta) data itself in a single place. It is also a comprehensive system, required to preserve the property right rather than the lighter touch voluntary systems that are present in a number of the recent EU proposals. Second, the paper argues that, similar to copyright, independent development of the same dataset should be allowable, combined with strong technical protocols on data covered by the property right[4]. Lenient grantback provisions and FRAND[5] terms are argued to be useful for licensing agreements to address both sequential innovation concerns and concerns about pricing of complementary inputs to follow-on innovations. Third, the paper argues that uptake of a newly created right will be challenging unless the "default" of trade secrecy protection (often combined with contractual provisions) is addressed: this alternative can be very attractive to data holders and is not necessarily the optimal choice for society and sets the paper apart from some others on the topic of data governance[6]. To generate uptake on a new right, the default rights regime must be considered and designed in tandem with any new right. Finally, one should consider whether the right should only remain in force if the data is "worked", in the

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

[4] See 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 as similar effect. Katznelson and Howells (2021) argue that narrow patents have the additional benefit of directing research optimally. This effect complements, and was not a focus of, the earlier work.

[5] Fair, Reasonable, and Non-Discriminatory

[6] See argument below. Trade secrecy is not fully analysed here. Cugno and Ottoz (2006) and Bessen (2005) 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, although 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. Anton and Yao (2004) find that patents and trade secrets are selected as protection for different types of innovation where choice is available, and Risch (2007) finds that, where costless to provide, allowing trade secrecy dominates not allowing it.

same way as gold claims had to be worked to retain protection to avoid purely defensive accumulation of rights.

Some systems contain elements from this list, but currently none of the proposed legislation takes explicit note of the default to ensure take-up of the right, nor as comprehensive a system of registration. Inclusion of these considerations into the list are, perhaps, the most notable departures from existing discussions.

In making these arguments this piece attempts to deal with economic, not legal issues of rights. Clearly, there must be some overlap with legal scholarship if we are to examine the creation of a right. A companion piece in this volume considers, then, the same type of question from a legal standpoint, leaving this piece with the liberty of a narrower, economic, approach.

The Lay of the Land

Focussing again on Europe, the set of legislation that impinges on data creation and use is as impressively complex as it is recent. Figure I includes selected legislation, including the upcoming Data Act, due to be announced at the end of 2021.

The effect of this plethora of legislation is difficult to gauge, as much is still at the proposal stage. We do, however, have a 2018 study[7] relating to one piece of legislation, the Database Directive, that outlines many of the concerns could be raised about the set of legislation in general.

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

This has some similarity with earlier studies of patent rights, which have found that innovators may use patent rights alongside other types of strategic protection (such as first mover advantages in marketing or learning by doing advantages) or other measures (secrecy, for example) to protect their innovations[8]. The menu of available protection measures for data is not the same as that earlier study of manufacturing innovations. For example, technological measures (eg, technological "locks" to prevent copying or using the data without authorisation) were not relevant to that earlier survey and strategic advantages such as learning by doing is not investigated in this survey. 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 the relevant choices are the same as for patentable technologies.

[7] 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.
[8] See the classic study by Cohen et al (2000) for details.

Second, the protection afforded by the directive is unclear. 40% of the respondents to the survey indicated that they had some level of uncertainty about whether they had infringed the right due to variations in the way it was implemented across European Union (EU) states. The confusion may simply result because of poor explanation of the right in the short amount of time since its creation. It may also be the result of poor awareness or from the fact that as a Directive individual member states have considerable leeway in how they interpret the requirements. Of course, a right that is poorly understood or simply not known is likely not to be particularly effective. The multiplication of rights that has occurred recently may result in further confusion generally.

Finally, respondents indicated that the *sui generis* right didn't necessarily contribute a great deal to their incentives to create new databases, amounting to less than 10% in all categories of innovation. This does not necessarily say that the directive is unimportant to innovation: the classic work of Cohen et al (2000) and others[9] suggest that patents are not heavily relied upon as a means of protection across a wide swathe of sectors but rather are most effective in a few concentrated portions of the economy such as pharmaceuticals and chemicals. Patent protection is, however, crucial where it is needed. If the same is true for data, we may well see relatively low percentages overall who rely on this protection instrument as an incentive to innovate without this "average" lack of take-up negating the fact that for some the protection might be both important and socially worthwhile. While the benefits evaluated across all sectors may be relatively low, the *sui generis* right seems to be associated with relatively low cost as well, so the net benefits may well be positive.

Overall, then, the current situation is one of a set of rights that may be applied to an innovation, perhaps together and perhaps as substitutes for each other, viewed as more, or less, effective and used in different combinations across different industries and governed by a complex and potentially confusing web of legislation, either now or in the near future. Designing a right that remains useful in the face of an existing web of protection instruments, is a challenge that shapes the conclusions that we reach, below.

What's the Rush?

There is no fully-functioning data collection system worldwide. Estimates are, however, available for the level and increase in data use. A recent EU report (EU, 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 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.

[9] See, for example, Moser (2012, 2013), Dosi et al (2006), Fontana et al (2016), and references therein.

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[10] that the global datasphere will grow to 51 zettabytes[11] by 2025.

It is not clear, then, that there is any pressing need for creating any further protections or incentives to generate more data. This is hardly a market that is struggling to develop under the current system. Still, robust growth does not necessarily mean socially optimal growth. For this, we would need to examine standard factors that drive a wedge between social and market optima. These include externalities, public good properties of data and other market failures associated with data, as well as the current system of regulation and protections that may impinge on market functioning. This could give us a hint of whether further protections were needed or whether the current system is adequate to achieve a likely social optimum.

On the question of externalities, Acemoglu et al (2020) points to the heterogeneity of data, with social value compared to private value likely to depend upon the specific type of data in question. Think of personal data, for example. Acemoglu and co-authors argue that here 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. On the other hand, 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.

In terms of the effect of data on market failures due to asymmetric information, increased data access may make markets work better or worse. Insurance markets typically suffer from asymmetric information. Asymmetric information can result in market failure and has sparked a large amount of regulation in insurance, but equally has led to extensive effort to detect the characteristics of policy-holders as part of pricing and enforcing contracts. To illustrate the scale of this problem, in 2019, fraudulent claims and dishonest applications amounted to 867000 cases in the United Kingdom (UK), with an annual value equalling approximately £1.2Bbn[12] in direct costs, but of course preventative measures add to this total. Obtaining better information in this market might well generate considerable savings. On the other hand, a market currently suffering incomplete but not necessarily asymmetric information could potentially 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[13].

---

[10] See Statista, https://www.statista.com/statistics/949144/worldwide-global-datasphere-real-time-data-annual-size/ Accessed 01/07/21.

[11] A zettabyte equals 1,000,000,000,000,000,000,000 or one sextillion bytes

[12] See Statista, https://www.statista.com/statistics/885271/fraudulent-insurance-claims-and-dishonest-applications-united-kingdom/) compared to a total contribution to the UK economy of £29.1Bbn, although this was in 2017, reported by a 2019 ABI https://www.abi.org.uk/globalassets/files/publications/public/key-facts/key_facts_2019_spread.pdf. US figures from the FBI indicate an estimate of $40Bbn per year in non-health insurance fraud. See https://www.fbi.gov/stats-services/publications/insurance-fraud 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/

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

All these remarks assume that the data that is added to the market is correct. If the added data is incorrect, then the anticipated gains in the functioning of the 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 a great deal to whether there is a social oversupply or a social undersupply. Furthermore, adding incomplete data may not move the market toward 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 is may also be based on false information, is less well-understood: the nature of the error in the data would affect the conclusion one would reach[14] and whether correcting the error would move the system toward or away from a social welfare improvement.

Finally, even if there are gains to a market due to increasing use of data, these need not be distributed to consumers. The use of consumer data in marketing can be a case in point. Consider, for example, data that allows better price discrimination through greater "targeting and customisation" for consumers. 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 the firm[15].

Overall, then, the argument for "more data" across the board is less compelling than the argument for certain types of data and the argument for more data access has some important caveats. In particular, there is likely a better argument for more data access to improve the overall efficiency of markets than to improve consumer welfare on its own.

We now turn to considering socially desirable data only, and how IPRs could be used to improve either the generation of data or data access.

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 is to focus on a right for the data itself, as is contemplated in the *sui*

[14] 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.
[15] Personalised markets may also modify bargaining power, possibly to the advantage of customers, but this is not investigated here.

generis right for databases.[16] We will focus on how IPRs work currently and then will see how data does or does not fit into this type of mould and what that implies for IPR design for data.

Stepping back first to consider why we have IPRs in the first place, innovation has been cited as key to economic growth[17] and promoting innovation is enshrined in key documents that establish the EU, the United States (US) and others[18]. This is despite the fact that economic theory is rather ambiguous about whether the market under or overprovides innovation compared to the social optimum. We might think that there is too little innovation, since innovators generate a positive externality. For example, when they create markets, inventors neither capture all the surplus generated (in standard posted pricing models), leaving some as consumer surplus, nor capture the benefits of knock on innovation conducted by rivals unless certain conditions on rights are fulfilled, as we discuss below. Indeed, where innovators are private firms or individuals, they can exert a negative externality when they develop their innovations to "steal business" from rivals. For example, a firm may innovate to keep an innovation from a rival, even if it has no intention of exercising the innovation (ie, a purely defensive use). Resolving which externality is larger becomes an empirical matter. Work by Bloom et al (2013) has shown, in fact, that the empirical evidence[19] points toward social under-provision of innovation, but this is an 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.

Still, intellectual property protection design is based on the assumption that some incentives need to be provided to increase innovation to the social optimum, in line with the empirical results quoted above. Most IPRs are designed to satisfy a "social contract" that balances the privilege of exclusivity, which establishes the incentive to innovate, with the obligation to diffuse or make public the protected innovation and the need to avoid excessive deadweight loss for the consuming population. Standard features, such as term and scope limits on IPRs balance the interests of innovator, follow on innovators, and the consuming public. Registration systems that explain and publicise the innovations where they are not publicised in the normal course of use promote the interests of follow on creators and even imitators. Facilitated diffusion, often undertaken with a registration or catalogue of past protected innovation improves information asymmetries by publicising the "guts" of innovations, facilitates further innovation that builds on previous advances by revealing best practice of the innovation; it directs innovative effort away from duplication and toward either completely new areas or design arounds that push the field forward.[20] Perhaps most importantly, a registration system facilitates market trade in the technologies themselves via licensing or other tools, which allows wider use, adaptation, and

[16] The control of those generating the data over 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. Here, a main concern is privacy regarding personal data. See Acquisti et al (2016) for a useful overview and references.
[17] See, for a useful overview of classic work from a variety of traditions, Verspagen (2006) and references therein.
[18] The US constitution also enshrines this (Section 8, see https://constitutioncenter.org/interactive-constitution/full-text), as does the Treaty of Rome in Europe (updated 2002, Title XVI, Article 157(1), OJ 325/103).
[19] That paper draws from US data.
[20] See Katznelson et al (2021, forthcoming JCLE) for a discussion of the difference between design around and pure imitation. See Scotchmer (1990, 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 and Lee, 2017).

improvement.[21] It does this by simply allowing potential licensing partners to become aware of each other and understand where there can be mutually beneficial trades. These design features of term limits, scope limits, and registration all potentially improve the progress of science, technology, and the creative arts.

Other policies clearly affect this balance of rights of different constituencies. 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, not in isolation, and achieving its aims of socially optimal innovation requires adjusting to changes in other parts of this web to the extent that the expected benefit stream from innovation changes[22]. Equally, the effectiveness of any one type of intellectual property protection depends on what the alternatives to this protection are. If technical protection, such as digital watermarking, or effective or "factual" trade secrecy[23] can be used to protect an intellectual asset well, then it may be used in place of or in addition to intellectual property protection. Clearly, if other methods substitute for IPRs, they cannot operate to adjust market behaviour. Hence, the innovative incentives of society as a whole depend on a system of rights that go well beyond traditional IPRs.

Overall, then, IPRs both incentivise production of innovations and provide an important diffusion function, which promotes further innovation. Access is an integral part and a *quid pro quo* of protection. One should consider both of these aspects when considering if an IPR can be designed to improve social welfare derived from data.

Gold is not Intellectual Property. Is Data Either?

Gold is property, but it isn't 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. A single data point also may not reflect much human creativity, but instead may simply be "found" in the course of undertaking normal business activity. At the same time, the EU has created an IPR-like *sui generis* right to protect databases based on the idea that once data is assembled into a database, it is no longer simply "found" but instead is curated. This step requires enough ingenuity and potentially monetary investment that it is worth protecting in order to ensure that the creator is compensated for that investment.

The paper is not addressing here whether (processed) data is eligible for a traditional intellectual property right. That question is large enough to be beyond the scope of this short paper. Instead, we look at how data differs from IPR-eligible innovations and, taking IPR design as a starting point, think

[21] Technology Transfer Guidelines in the EU and US summarise the large literature on licensing by noting that in general it has positive welfare consequences, although with notable exceptions. For the EU, see Guidelines (2014) or TTBER (2014) and for the US see US DOJ/FTC (2017).
[22] This argument is developed in Regibeau and Rockett (2007). See also Gilbert (2020).
[23] See Sandeen and Aplin (2021) for a discussion of "factual secrecy", which captures the idea of effective trade secrecy independent of any necessary legal right.

about how those differences impact the type of right one might wish to design. In other words, we will take the approach taken during the Gold Rush of starting with a type of law that may not be a perfect fit but has some useful features, and then will build from there toward something that might fit the purpose at hand.

Data – in the sense that we are using it in this paper as processed data - is an input to socially valuable innovation, without intrinsic value in itself. It is a research tool and may be a necessary one: without it 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, it can be the first step in a process. Indeed, just because a research tool has no end value in itself does not mean that it should be ignored: the benefits from its offspring can be very large for economic growth among other things.

This role of data in a sequential stream of innovation is a form of complementarity: as a research tool, data is complementary to its own offspring. Data is also complementary in other ways, however. Within the realm of data itself, as opposed to the innovations that might spring from its use, datasets may be complementary to each other. Dataset A is not necessary to collect dataset B in most cases; however, using datasets A and B together can produce insights that were not possible with either type of data on their own.

Data can be an innovation market gatekeeper[24] in either a sequential innovation framework or in a framework where is it a necessary element of a larger data gathering exercise. For example, it may not be possible either to evaluate the quality of existing disease treatments or improve on those treatments without data on patient performance after exposure to a particular therapy. As such, data may be a gatekeeper to innovation if it is closely held or non-replicable from other data sources. Broader protection, allowing the research tool to capture some of the profit from follow-on innovations has been proposed as a way to solve the problem of incentivising the creation of research tools and follow-ons that build on them.[25] At the same time, insights into better treatments for patients may not be possible without both a dataset on patient outcomes and information on the larger demographic and socioeconomic status of those patients and others who share similar features. In this case, both datasets are necessary to generate treatment insights. We know from other work, largely focussing on standard-setting, that in cases where several necessary pieces of a system must fit together to generate an innovation, then hold up and excessive pricing become potential hurdles in accessing those pieces.[26]

These twin difficulties may generate different solutions in rights design. If there is sequential complementarity, then broad rights for the research tool have been recommended in order to allow the

[24] Here, we do not link gatekeeping to product market share, but rather point to its role in facilitating innovation within innovation markets. In this sense, it is closer to the role of data as an innovation technology as in Prufer and Schottmuller (2021). Here, however, we are talking about processed data, requiring investment to create, which is one point of distinction between our work from theirs.
[25] For this to work smoothly, *ex ante* licensing must be both possible and legal. See Green and Scotchmer (1995), but also see also Ambashi et al (2020) for a discussion of grantbacks to address this issue.
[26] 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.

tool to capture its own positive externality. If there is complementarity between datasets, then allowing datasets to be pooled and licensed under FRAND commitments may be an appropriate solution[27].

Where broad protection is used, the sequential innovation literature indicates that the research tool will be shared by relying on Coasian logic[28]. It is not clear that this is what we observe: inadequate diffusion tends to be a more general complaint than inadequate data generation. One possibility is that data is not protected strongly enough so that the rights regime that makes the Coase Theorem work well does not apply. Another is that data is protected strongly enough, but the underlying assumptions of the sequential innovation literature that generate conclusions about the optimality of strong rights do not hold in the case of data.

While data may be relatively easily removed from a place of work by an employee, as it is highly transportable, it is also the case that contractual obligations can restrict this behaviour. We will set aside the first consideration and instead focus on the second possibility. Here, there is a potential difference between the sequential innovation's focus on the case where data-holder must pair with other independent entities to generate an innovation. If a single firm cannot both obtain the data and innovate then we can think of ideas as "scarce" in the sense that the research tool maker may not also have the idea of the follow on innovation[29]. This scarcity and attendant need to pair with others may be true in some cases, but it is not uncommon at all for a data-holder to be an innovator in its own right, using 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 to innovate. Hence, it is not clear that the case of "scarcity" applies to the datasphere.

A second feature of data, other than the potential for the data holder to also be innovator, is that data often is held as a trade secret, as we have seen in the survey data presented above in figure 2. Indeed, 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. Instead, data can be used within the confines of the body that created the dataset to create further innovation in the same way as a process innovation may be used within a firm to generate further innovation and remain protected with trade secrecy. Indeed, perhaps analogously, one observes lower reliance on patents for process innovations where a trade secrecy option is viable.[30] [31]

The fact that secrecy is a viable and attractive alternative for protecting data has two implications where data are also a research tool in a sequential innovation process. First, if data are relatively easy to hide

[27] The appropriate design of the patent pool is discussed in 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.
[28] For example, see Green and Scotchmer (1995).
[29] Scotchmer (2004) expands on the idea of ideas being scarce.
[30] Cohen et al (2000) report that process innovations tend to be protected more frequently with trade secrecy than product innovations.
[31] 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 provisions may also be favoured protection measures 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.

whilst being used to create value, then uptake of any new right associated with diffusion is by no means guaranteed: if trade secrecy – factual or legal - is enough for all the exclusivity the creator needs then an additional right could be superfluous. This suggests, that the incentive to take up the right may be far less obvious than it is for patented, copyrighted, or trademarked material.

Trade secrecy may not, however, obtain the careful balance of interests that other IPRs are designed for. While trade secrecy may implement an independent invention system, as advocated by Maurer and Scotchmer (2002), it does not have the positive effects of directing innovation as does the patent system. This function comes through the patent disclosure, which can be used to identify business opportunities (Lee and Lee, 2017), learn about competition and so optimally adjust their own research plans (Oppenheim, 1998), or use the patent as a springboard to socially beneficial "invent arounds" (Katznelson and Howells, 2021). In other words, the diffusion aspect of the patent register is important, and is not part of trade secrecy. The lack of disclosure creates, then, a potential loss for society.

Second, data may not automatically enter the public domain when used but instead may need to be placed there if the data's existence is to be widely known. While contractual means, such a licensing, can be used to trade data and so generate social value from a stream of innovation if the data are held as a trade secret, without wide knowledge about the data's existence or nature, anonymous or unknown parties will not be aware enough of the data to contact the data-holder to arrange access. A central clearing house could do a much better job of this. Such a clearing house would widen the categories of data that would be registered from the restricted categories currently covered.

Hence, secrecy does not perform the useful purpose of wide and diffusion to parties that are potentially unknown to the innovator of, for example, a patent system or a trademark registry. Furthermore, trade secrecy combined with the ability of the data holder to also – eventually -generate many of the innovations flowing from the data can mean that data is released too slowly and innovations follow with more delay than the social optimum[32].

<u>If Data is Data, not Gold, What Does It Mean for Property Rights?</u>

One can make several comments about features of a protection system for data based on these observations. These distinctive features can form the basis for designing a right for this new area.

First, diffusion via a market cannot develop if agents do not know with whom they should partner[33]. Even if one continues to rely on trade secrecy, a registration system that creates a central repository

[32] See Matutes et al (1996) for background on this argument. That paper 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.

[33] Using a market to govern diffusion allows data that is 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.

for data holdings, potentially on a worldwide basis, could promote socially beneficial trading to develop across entities that are not known to each other *ex ante*. A registration system would be one way to provide one stop shopping for data, allowing for improved information to those who could benefit from data but who might be unexpected users. Making such a system worldwide allows for the best opportunities for matching up complementary data. While various worldwide repositories exist for limited types of data, there is no agreed protocol for worldwide holdings, access, or a single place to find all data[34].

The difficulty of designing such a repository is to specify the data in sufficient detail for potential partners to know what is there, but with enough protection that the data are not disclosed so as to protect creators. Patent repositories that are created to promote further innovation, disclose so fully that – if applied to data, would put exclusivity in jeopardy; however, it is not clear that facilitating a market for data requires access to the content of datasets. There must be some incentive to enter the repository and full disclosure could remove this. Short of listing the data itself, cataloguing systems using metadata to classify and cross reference the data down to relatively specific subject matter in the same way as technological categories are used in patent systems. This in no way compromises the contents and allows partnerships to develop to obtain "systems" of data necessary to resolve identified questions. Of course, those who wish to make the full dataset open access can do so and a comprehensive repository also helps them[35].

Independent invention is normally not sanctioned for protection for patentable materials but is for copyrighted materials, as long as those materials are not copied. The (innovation) gatekeeping capability of data often derives from protection of the data source, for example a large customer base that is kept relatively exclusive by means of an underlying draw from other business activities. This customer base can be well protected by strategic or behavioural processes, such as network externalities or the ingenuity of the platform curator, rather than any type of legal protection. If the balance of challenges in the datasphere is access rather than the incentive for creation, then there is an argument for an independent invention protection for data, where anyone with the ingenuity to generate the same data would be protected as long as there is no direct evidence of copying.

The accuracy of a dataset must be known to the extent possible to make the data useful to those who wish to use it for further innovation. Curation according to standard protocols is already recommended in the EU for certain types of data, so this is a matter of extending these protocols to a wider variety of data. Both reliable and unreliable data undoubtedly have their place: those who study fake news would likely benefit from invalid or otherwise completely made-up data. Still, being able to judge the quality of the data and provide a label to identify this quality could govern data's appropriate uses.

[34] Current 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.trilegal.com/index.php/publications/analysis/revised-report-by-the-committee-of-experts-on-non-personal-data-governance?utm_source=Mondaq&utm_medium=syndication&utm_campaign=LinkedIn-integration

[35] 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.

Trading these data would occur by license. Aside from open access licensing and ensuring that licensing occurs within other laws (such as privacy laws), private licensing guidelines can allow for measures allowing the "sequential" and "horizontal" complementarities to be captured, at least in part. Grantback provisions are a ready tool to allow for data-makers to benefit follow on uses. Grantbacks (not necessarily free) can reduce the problem of hoarding data for one's own use rather than allowing the innovations to be generated and then capturing some of the benefit. FRAND provisions are a standard tool to control pricing of complementary inputs to innovation, although this system is not without its critics.

Third, a standard means of balancing user and maker rights is to impose term limits on exclusivity. There is no reason why this would not continue to be true for data, meaning that the registered data would need to become open source at the end of its term of protection. In other words, using the spectrum of openness illustrated by the Open Data Institute (figure 3, below), the access rights at the end of the term of protection would need to be to the right (and perhaps the far right) of that spectrum.

Of course, to get uptake of any new right that is term-limited, one would need to make the limited right more attractive than the alternative of secrecy. Hence, the "default" protection mechanism needs attention, not just the design of the target right. When considering whether society would benefit more from protection coupled with revelation or protection via trade secrecy, it is important to consider that if trade secrecy is selected there is a potential loss to society. This is because trade secrecy does not necessarily balance maker and user rights in any kind of socially optimal way. For example, the *quid pro quo* for patent protection is revelation of the innovation and this revelation as well as the term of protection are centrally set so that there is the option of the governing body to establish these in a way that maximises social welfare. Complementary to this "right to reveal" as part of the patent system trade secrecy provides protection without revelation[36]. The choice of which type of protection to use, and the selection of an effective term of trade secrecy, including when and to whom to reveal information, normally is decided by the agent holding the secret and so would generally occur at the point where the agent's best interests are served, not necessarily society's interests as a whole.

Hence, from society's point of view, there is a good reason to incentivise data-holders to take up a right that is coupled with revelation and open access after a term that is centrally set by a governing body maximising social, not individual, benefit. A default of open access for non-right holders or compulsory licensing under some conditions performs this and, as with some of the other points in this article, has been explored in the context of certain types of data. Registration and abiding by a licensing and curation protocol would then be required to retain rights. A modification of this idea is a system that would require that the data be "worked" to retain rights as, indeed, claims were required to remain worked to keep right active during the Gold Rush and as trademark coverage requires in some jurisdictions. This avoids collecting data simply to keep it from others. The point is that without attention

[36] See Paine (1991).

to the default, uptake may well not occur at all. Indeed, this has been observed for the *sui generis* database protection.

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 equal granularity in the data registration system. Further, to the extent that artificial intelligence systems may shift the balance of research toward 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 future.

The differences between this set of suggestions for data rights and a standard IPR right are in degree, rather than in nature. 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 taken in conjunction with work on designing a new property right.

Conclusions

Data are data, and gold is gold. They aren't the same and shouldn't be protected in the same way. At the same time, the protection of gold claims in the Gold Rush of the mid-19th century in the US faced similar problems of creating rights to a newly-discovered resource. In line with current models available, such as the *sui generis* right in the EU, and the exercise that is attempted here, previous legislation and experience was tapped, with adjustments to accommodate the new circumstances to try to create a right fit for purpose.

While data may not be classic intellectual property, at the same time data share some features of intangible or intellectual assets that would tend to be protectible through IPRs: externalities, complementarities with both sequential innovations and other data, and non-rivalry. This suggests that intellectual property protection may be an appropriate model to start from in designing data protection rights, even if a data right does not end up as a full IPR.

The current EU system, let alone the world system, of protecting data is fractured. This is both confusing for users and unhelpful, given that data sharing can generate valuable gains and that confusion can hamper uptake. Some steps have been taken to design IPR-like rights for data, including in the form of the *sui generis* right. These systems, in the view of this author, could be more effective if they were

extended, emphasised registration and cataloguing, and if they were undertaken in conjunction with a review of other methods currently used to protect data. If these alternative methods of protection are too favourable, particularly if they allow for secrecy to be a main tool of protection, then there is little incentive to take up any newly created right. Excessive use of secrecy may not promote the delicate balance of constituent interests that other IPRs allow, nor promote sufficient diffusion for a vibrant data markets to develop. This could be a loss for society, slowing innovation. The point of this paper is to sketch some key elements of a data right. While some have been outlined before, the point that attention needs to be given to default rights has not.

References

Acemoglu, D., Makhdoumi, A, Malekian, A, and Ozdaglar, A (2020) "Too Much Data: Prices and Inefficiencies in Data Markets" available at . https://economics.harvard.edu/files/economics/files/acemoglu_spring_2020.pdf [accessed November 1, 2021]

Acquisti, A., Taylor, C., and Wagman, L. (2016) "The Economics of Privacy" *Journal of Economic Literature* 54(2), 442-492.

Ambashi, M., Regibeau, P. and Rockett, K. (2019) "Grantbacks, Territorial Restraints, and Innovation" *International Journal of Industrial Organization* 67 (December).

Anton, J., and Yao, D. (1994) "Expropriation of Inventions: Appropriable Rents in the Absence of Property Rights" *American Economic Review* 84(1), 190-209.

Anton, J., and 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" available at https://www.abi.org.uk/globalassets/files/publications/public/key-facts/key_facts_2019_spread.pdf [accessed November 1, 2021].

Bessen, J. (2005) "Patents and the Diffusion of Technical Information" *Economics Letters* 86(1), 121-128.

Bloom, N, Schankerman, M, and Van Reenen, J. (2013) "Identifying Technology Spillovers and Product Market Rivarly" *Econometrica* 81(4), 1347-1393.

Choi, J. P. (2002) "A Dynamic Analysis of Licensing: The 'Boomerang' Effect and Grant-back Clauses" *International Economic Review* 43(3), 803-829.

Clay, K, and Wright, G. (2005) "Order without Law? Property Rights during the California Gold Rush" *Explorations in Economic History* 42(2), 155-183.

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. and Ottoz, E. (2006) "Trade Secret vs Broad Patent: The Role of Licensing" *Review of Law and Economics* 2(2), 209-221.

European Commission (2014) "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 (2014) "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.

Deloitte (2018) "Data is the New Gold: The Future of Real Estate Service Providers" available at: https://www2.deloitte.com/global/en/pages/real-estate/articles/future-real-estate-data-new-gold.html [accessed 1 November, 2021]

US Department of Justice and Federal Trade Commission (2017) "Antitrust Guidelines for the Licensing of Intellectual Property" available at: https://www.ftc.gov/system/files/documents/public_statements/1049793/ip_guidelines_2017.pdf (accessed November 21, 2021).

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.

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. Available at: https://op.europa.eu/en/publication-detail/-/publication/5e9c7a51-597c-11e8-ab41-01aa75ed71a1/language-en [accessed November 1, 2021].

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 available at: https://op.europa.eu/en/publication-detail/-/publication/9fb0599f-c18f-11ea-b3a4-01aa75ed71a1/language-en [accessed November 1, 2021].

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

European Union, *Treaty Establishing the European Community (Consolidated Version), Rome Treaty*, 25 March 1957, available at: https://www.refworld.org/docid/3ae6b39c0.html [accessed 1 November 2021], updated 2002 as Consolidated Version of the Treaty Establishing the European Community, OJ. 325/35, 24/12/2002.

Federal Bureau of Investigation Reports and Publications, (undated) "Insurance Fraud" available at: https://www.fbi.gov/stats-services/publications/insurance-fraud [accessed 1 November, 2021]

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.

Forbes, Africa (2019) "Data is the New Gold", available at: https://www.forbesafrica.com/technology/2019/07/18/data-is-the-new-gold/ [accessed 1 November, 2021]

Fosfuri, A. and Ronde, T. (2004) "High Tech Clusters, Technology Spillovers, and Trade Secret Laws" *International Journal of Industrial Organization* 22(1), 45-65.

Frankel, J. and Rockett, K. (1988) "International Macroeconomic Policy Coordination when Policy-Makers Disagree on the Model" *American Economic Review* 78(3), 318-340.

Gallini, N, and Wright, B. (1990) "Technology Transfer under Asymmetric Information" *Rand Journal of Economics* 21(1), 147-160.

Gilbert, R. (2020) *Innovation Matters: Competition Policy for the High Technology Economy* Cambridge, MA: MIT Press.

Gilbert, R. and Newbery, D. (1982) "Preemptive Patenting and the Persistence of Monopoly" *American Economic Review* 72(3), 514-526.

Goshen, Z., and Parchomovsky, G. (2001) "On Insider Trading, Markets and 'Negative' Property Rights in Information" *Virginia Law Review* 87(7), 1229-1277.

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

Green, J., and Scotchmer, S. (1995) "On the Division of Profit in Sequential Innovation" *Rand Journal of Economics* 26(1), 22-33.

Halloran, D. and D'Souza, F. (2020) "Data is the New Gold: This is How it can Benefit Everyone – While Harming No One" *World Economic Forum* available at: https://www.weforum.org/agenda/2020/07/new-paradigm-business-data-digital-economy-benefits-privacy-digitalization/ [accessed 1 November, 2021]

Katznelson, R. and 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,* available at: https://doi.org/10.1093/joclec/nhab010.

Lee, M. and 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.

Lemley, M. and Shapiro, C. (2007) “Patent Holdup and Royalty Stacking,” *Texas Law Review* 85, 1991-2049.

Lerner, J. and Tirole, J. (2004) “Efficient Patent Pools,” *American Economic Review* 94(3), 691-711.

Matutes, C., Regibeau, P. and Rockett, K. (1996) “Optimal Patent Design and the Diffusion of Innovations” *Rand Journal of Economics* 27(1), 60-83.

Maurer, S. and Scotchmer, S. (2002) “The Independent Invention Defense in Intellectual Property” *Economica* 69, 535-547.

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, 23–44.

Paine, L. (1991) “Trade Secrets and the Justification of Intellectual Property: A Comment on Hettinger”, *Philosophy & Public Affairs*, 247-263.

Prufer, J. and Schottmuller, C. (2021) “Competing with Big Data” *Journal of Industrial Economics*, forthcoming.

Regibeau, P., and Rockett, K. (2007) “The Relationship between Intellectual Property Law and Competition Law: An Economic Approach” in Steven Anderman, Editor, *The Interface between Intellectual Property Rights and Competition Law*. Cambridge: Cambridge University Press. Chapter 10, 505-552.

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., and Aplin, T. (2021) "Trade Secrecy, Factual Secrecy and the Hype Surrounding AI," in Ryan Abott, Editor, *Research Handbook on Intellectual Property and Artificial Intelligence.* Cheltenham: Edward Elgar. Forthcoming.

Scotchmer, S. (1991) “Standing on the Shoulders of Giants: Cumulative Research and Patent Law” *Journal of Economic Perspectives* 5(1), 29-41.

Scotchmer, S. (2004) *Innovation and Incentives*. Cambridge, MA: MIT Press.

Scotchmer, S. and Green, J. (1990) “Novelty and Disclosure in Patent Law” *Rand Journal of Economics* 21(1), 131-146.

Spulber, D. (2017) “Complementary Monopolies and Bargaining” *Journal of Law and Economics* 60(1) 29-74.

Statista (2021) “Number of Fraudulent Insurance Claims and Dishonest Applications Detected in the United Kingdom (UK) from 2017 to 2019,” https://www.statista.com/statistics/885271/fraudulent-insurance-claims-and-dishonest-applications-united-kingdom/ [accessed 1 November, 2021]

Statista (2021) “Annual Size of Real Time Data in the Global Datasphere from 2010 to 2025” available at: https://www.statista.com/statistics/949144/worldwide-global-datasphere-real-time-data-annual-size/ [accessed 1 November, 2021]

The Constitution of the United States, available at: https://constitutioncenter.org/interactive-constitution/full-text [accessed 1 November 2021]

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

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.

Verspagen, B. (2006) “Innovation and Economic Growth” *Oxford Handbook of Innovation*. J. Fagerberg and D. Mowery, Eds. Oxford: Oxford University Press.

World Intellectual Property Organisation, WIPO (2002) “Summary on Existing Legislation Concerning Intellectual Property in Non-Original Databases” SCCR/8/3.

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

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 (under revision) – 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 (Proposal – 2020) – 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 (Proposal – 2020) – various hosts, intermediaries, and platforms. Requires transparency and compliance with fundamental rights and various codes of conduct.

Data Governance Act (Proposal – 2020) – improve trust through voluntary labelling, data-sharing structures including interoperability (esp. public sector data), horizontal board to oversee data governance.

Data Act (Proposal Q4 2021) – 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 2: 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"

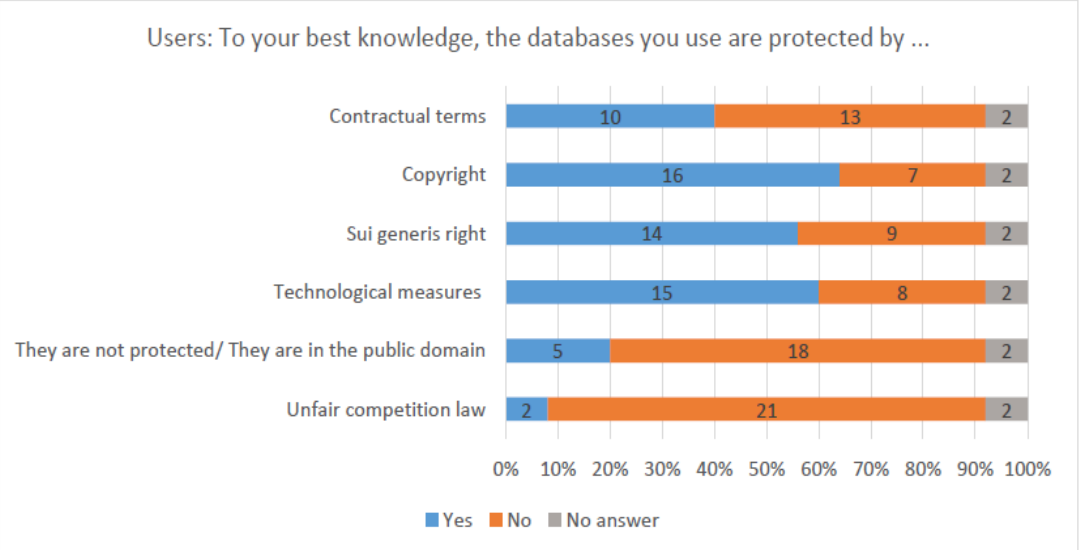

Figure 3: The Data Spectrum, from Open Data Institute, https://theodi.org/about-the-odi/the-data-spectrum/