Data as Collectively Generated Patterns: Making Sense of Data Ownership

Mathias Risse
Carr Center
Discussion Paper
Data as Collectively Generated Patterns:  
Making Sense of Data Ownership

Carr Center for Human Rights Policy  
Harvard Kennedy School, Harvard University  
April 26, 2021

Mathias Risse  
Faculty Director, Carr Center for Human Rights Policy  
Berthold Beitz Professor in Human Rights, Global Affairs and Philosophy

The views expressed in the Carr Center Discussion Paper Series are those of the author(s) and do not necessarily reflect those of the John F. Kennedy School of Government or of Harvard University. Faculty Research Working Papers have not undergone formal review and approval. Such papers are included in this series to elicit feedback and to encourage debate on important public policy challenges. Copyright belongs to the author(s). Papers may be downloaded for personal use only.
ABSTRACT: Data ownership is power. Who should hold that power? How should data be owned? The importance of data ownership explains why it has been analogized to other domains where ownership is better understood. Several data-as proposals are on the table: data as oil, as intellectual property, as personhood, as salvage, as data as labor, etc. Here I propose another way of thinking about data. Like the others, my view characterizes data in ways that make them accessible to ownership considerations and can be expressed as a data-as view: data as collectively generated patterns. Unlike the alternatives, data as collectively generated patterns does not create any equivalence with another domain where ownership is already well-understood. It reveals how ownership considerations enter, but we must explore afresh how they do. Accordingly, I propose a way for ownership considerations to bear on data once we understand them that way. And if we did understand them that way, the internet should presumably be designed very differently from what we have now. The themes in this paper require more elaboration than what can be done here, so some follow-up work is needed.

1. Introduction

We live in the age of big data. Every day, electronic devices - from laptops or smart phones to GPS’s, learning thermostats and virtual assistants - collect myriads upon myriads of data about all manner of human activity. The relevance of data has become so stark that it has been described in terms of some striking superlatives. For one thing, the collection of data has reached such dramatic dimensions that Shoshana Zuboff has coined the term “surveillance capitalism” to capture its importance for how companies prevail on markets. Alternatively, with an eye on emphasizing the harmful effects of data collection, Bruce Schneier has long insisted that data is the exhaust, and thus the pollution, of the information age, comparable to what air pollution was to the industrial age. And in a somewhat more reverential if perhaps facetious spirit Yuval Harari and others talk about “dataism” to capture the idea that information flow is the supreme value of the current age.

Access to data allows for detection of behavioral patterns that enable predictions for what individuals will do next, or what will happen to them. In the aggregate, such predictions anticipate societal trends. Whoever can make such predictions can benefit from them, or even redirect behavior. For example, Amazon Web Services controls vast arrays of cyberspace, through which data collection at a breathtaking scale becomes possible, which in turn could be used to guide customers on the Amazon website. Or data could be used to start entirely new trends. The more data there are, the more accurate such efforts become, and the harder it is for competitors to outperform successful efforts. Accordingly, data ownership is genuine power. Who should hold that power? How should data be owned?

The importance of data ownership explains why it has been analogized to other domains where ownership is better understood. Several data-as proposals are on the table: data as oil, as intellectual property, as personhood, as salvage, and Data as Labor. Each time the point is to say that data essentially is, or is relevantly like, something else about which we already comprehend how it can be owned. More specifically, data-as proposals do three things. First of all, they make a suggestion for what it is about data that makes them valuable; secondly, by identifying how data are valuable the proposal points to who should own the data; and thirdly, the proposal points to how data should be owned, to the kind of rights involved. For instance, according to Data as Labor, data are valuable because they amount to labor provided by specific people; therefore, then, the data provided should be owned by the person who provided that labor; and the data should be owned the way labor normally is owned in that society. For practical purposes that third question is most relevant.

My goal is to propose another way of thinking about data. Like the others, my view characterizes data in ways that make them accessible to ownership considerations and can be expressed as a data-as view: Data as Collectively Generated Patterns. However, unlike its alternatives, Data as Collectively Generated Patterns does not create any equivalence with another domain where ownership is already well understood. Nor, therefore, does Data as Collectively Generated Patterns straightforwardly reveal who should own the data and how they should be owned. Such work needs to be done separately. For the various tasks involved here I draw on my earlier work on intellectual property.

“Data” is the Latin word for “given.” In the sciences, data are the observations the theory must explain. In the digital world, data is anything electronic devices have recorded. In isolation almost all such recordings are without value. I say “almost all” by way

1. In 2017, 2.5 quintillion bytes of data (so that is 2.5 followed by 18 zeroes) were produced every day on average; see this 2017 DOMO infographic https://perma.cc/FY2R-RUG5. It is estimated that by 2020 each person will be “generating” 102 MB per minute on average; see this 2018 DOMO infographic: https://perma.cc/RY-KCXR. I am grateful to audiences at Harvard’s Center for Research on Computation and Society and at MIT’s philosophy research colloquium for helpful discussion. I also owe much inspiration from a discussion in the graduate fellow colloquium at the Edmond J. Safra Center for Ethics, which I co-direct with Meira Levinson. In February 2019, the colloquium discussed a chapter by Elettra Bietti that explored several aforementioned data-as proposals. It was as part of my own reflections on that material that I developed the particular way of thinking about data ownership that I present here and connected it to my earlier work on Grotius and intellectual property.

2. Zuboff, The Age of Surveillance Capitalism. See also Schneier, Data and Goliath.


5. I switch to capital initials because these are names of positions with particular contents.

—but makes them an

whether they succeed for the

that is not my concern here, but these considerations apply straightforwardly to intellectual property. In this way, much less extensive ownership rights to ideas are justifiable than in the

extent, it is because they reflect patterns of behavior that generate predictions. Normally these will be patterns formed through behavior of numerous people, which gives rise to the phrase “data as collectively generated patterns” as a way of describing what it is about data that makes them economically relevant and thus makes regulation of their ownership politically important.

Current practice is that whoever collects data controls them, and by letting this status quo prevail we might already have moved a long way towards creation of an ownership arrangement for the future with enormous economic and political implications for possibilities of human living arrangements. But there are substantial concerns about this, both at the macro-level (in terms of its future economic and political consequences) and at the micro-level, often formulated in terms of “privacy,” one way or another. Therefore, and since data ownership matters enormously, we have good reason to treat questions of data ownership as genuinely open at this stage. While for law and public policy some continuity with the status quo has distinct advantages to avoid objections of the “that-is-unrealistic” sort, this is still a good time for philosophers to bring to bear even unorthodox thinking on the matter. I proceed in that spirit.

A classic work in property theory that is of surprising value here is Hugo Grotius’s Mare Liberum, Free Sea, first published in 1609. Grotius’s topic is ownership of the seas in the context of the European expansionism of the 17th century. For various reasons Grotius held that the seas should remain unappropriated. That view has, by and large, prevailed over time and is reflected in international law. But it was distinctly controversial in Grotius’s time. After all, if far-flung lands could be claimed, why would that not be true of waterways one must traverse to reach them?

Mare Liberum offers a range of pro-tanto considerations against appropriation in any domain. Whether they succeed for the seas is not my concern here, but these considerations apply straightforwardly to intellectual property. In this way, much less extensive ownership rights to ideas are justifiable than in the dominant approach to intellectual property that goes back to John Locke. These Grotian considerations also apply to collectively generated patterns. Such patterns are not ideas but are relevantly like ideas. That makes regulation of their ownership susceptible to the considerations in Mare Liberum — considerations that should have, but never have, found their way into the regulation of ownership of ideas. As a result, ownership rights to data can only be justified to a much more limited extent than other approaches allow or is reflected in current practice. It would be immensely desirable for a new kind of internet to be designed that reflects a revised ownership structure.

Section 2 explores why ownership considerations matter in this context to begin with. Section 3 discusses data ownership. Section 4 introduces and rebuts various alternative data-as proposals and then also presents Data as Collectively Generated Patterns. Sections 5 and 6 introduce the Grotian approach to intellectual property by way of contrast with Locke’s and transfer those ideas to the context of data ownership. Section 7 reconnects to the discussion of Data as Collectively Generated Patterns, and section 8 concludes.

2. Why Ownership?

One objection to this whole endeavor is that ownership may not be the right way to address concerns about data use at all. By way of validating the present line of inquiry I attend to that objection right away. The objection can be formulated in two ways. The first is to point out that concerns about data ownership often involve privacy, and should be assessed directly, in ways that neither enlist nor benefit from considerations of ownership. Ownership language is redundant. The second way of pressing this objection is that, while ownership considerations might matter, nothing much hangs on how we determine the ownership status of data. Regardless of whether data can be privately or publicly owned, concerns expressed in terms of privacy would compete with them and entail the same conclusions.

Consider some cases to develop these matters:*

Case 1: Data analysis reveals that a certain pattern of typing on keyboards indicates the onset of a debilitating muscle disease. Insurance companies pay for such information because it allows them to reject potential clients or adjust premiums.

Case 2: Data analysis reveals that someone’s reading tastes make it likely that she would be receptive to advertising for a new book on internet privacy. Internet sellers that carry that title pay for that information to improve marketing of their product.

Case 3: Data analysis renders it likely that somebody is the kind of voter who could be persuaded to favor a particular candidate in response to a certain type of input (regardless of whether that input is accurate, which she would not be able to verify). Political campaigns pay for such information because it increases their chances at winning elections.

In such cases, so objectors could insist, deliberation about the appropriateness of using data in such ways could proceed without consideration of ownership. In Case 1, that debate revolves around the kind of intrusion constituted by a transfer of information from one context (being active on a keyboard) to an entirely different context. Such analogies have limitations, but they provide the basic structure for what I call a micro-level consideration of ownership.


8. Versions of these examples were brought up by Lawrence Lessig in the discussion of the Bietti paper in February 2019 (see footnote 1). I may not capture them accurately or use them in a somewhat different way. But the point was to probe the usefulness of ownership considerations in this context.
one (medical assessments). In Case 2, a commodification of day-to-day behavior occurs that stands in need of justification. In Case 3, the discussion revolves around the permissibility of deploying particular means in the process of opinion formation.

But upon reflection, ownership considerations do enter. In Case 1, one may be appalled at the intrusion, but matters are not entirely simple. Any conclusion we may reach involves pondering multiple considerations. One concerns what obligations insurance companies have, or how they should be regulated. Sharing insurance also means sharing risk: risk-sharing is the point of insurance. But if an insurance company knows that somebody is at a higher risk it would thereby also impose higher average costs on others who are in no way connected to the fate of the high-risk client. These others may well reasonably complain if information about risk profiles is available to the insurance but not used to assess premiums. But if the case is not free of controversy, ownership considerations might well enter by way of making clear that the decision needs to be made by the person who owns the data, giving suitable play to competing moral considerations.

In Case 2, one could argue that the kind of commodification involved is rather trivial. In such cases, commodification would likely be welcomed, or in any event tolerated, by potential buyers. Targeted advertising conversely is a way of protecting people from a flood of pointless marketing. In that case again a judgement needs to be made. Per default it should be made by a person who has an ownership claim to the data, which would again reveal a way in which ownership considerations matter after all. Finally, in Case 3 one could argue that there will be reasonable disagreement about scope and limits of legitimate means of persuasion, which again would point to a role for ownership considerations.

In other words, the response to all these cases is that ownership may not be the whole story: even within the domain of rights other types of right may enter, such as some type of liberty- or personality-related rights. But saying that much is consistent with ownership being one of the considerations that matter, and in each case, there is some reason to say that that is indeed so.

Another way of making the point of the relevance of ownership consideration is to resist the importance of privacy considerations. The term "privacy" merely points to a relevant distinction between something that should be left to individual decision-making, be hidden from eyes and ears of others or not be subject to particular scrutiny or assessments, on the one hand, and other matters that should not be set aside in such ways. "Privacy" or "the private sphere" is not independently understood. Therefore, independently secured considerations of privacy cannot readily be enlisted to limit the relevance of ownership considerations. Instead, we need arguments to establish a particular way of drawing such distinctions. Ownership considerations might well help with that.

Generally, the point of ownership is to generate a set of claims, liberties and powers to do with certain things as one pleases within certain legal and arguably also moral constraints. In none of these three cases do ownership considerations by themselves settle the matter conclusively. But nor do such considerations just not matter all. And as the old adage has it, possession is nine-tenths of the law. The bundle of claims, liberties and powers that constitute ownership is, for better or worse, a key component of our social lives because it helps with assignments of responsibilities and shapes expectations.

The bundle of claims, liberties and powers that constitute ownership is, for better or worse, a key component of our social lives because it helps with assignments of responsibilities and shapes expectations. Reflection on data ownership is reflection on how to make room for data in a world that works that way. That is the general point of bringing up ownership considerations in this context.10

What about the point that it would not matter how the ownership status was resolved? This would be the case if consideration of private ownership in combination with moral constraints on such ownership would always lead to the same conclusions as considerations of public ownership of sorts. If such an equivalence could be shown (for suitable understandings of private and public ownership), that would be remarkable. But in any event the task at hand is to clarify the nature of ownership here to begin with, and only then to wonder how this might be different from or equivalent to other ways ownership questions can be resolved.

3. Data Ownership

Literally, "data" is something that is, in some sense, given. In the digital world, data is anything electronic devices have recorded. Wikipedia defines data (in the context of computing) as "any sequence of one or more symbols given meaning by specific act(s) of interpretation." Data is not the same as information. "Information" goes back to the Latin for "giving shape to." In that spirit information is something that resolves uncertainty or otherwise offers some news value to the recipient. Data may be information if it has such value. For Claude Shannon, the founder of information theory, the point of information is the degree to which it surprises, which is context-specific. This feature sets that notion apart from both data and content (the latter being a neutral way of referring to what data contain).

In certain contexts, "data," or some type of data, is defined more specifically, as for instance "personal data" is in the European Union’s General Data Protection Regulation (GDPR). Under Article 4 (1) of the GDPR, personal data is any information relating to an identified or identifiable natural person ("data subject"); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person.

9. On that point about the relevance of context, see Nissenbaum, Privacy in Context.

10. For the view that talk about property in the case of data is inappropriate altogether, see https://medium.com/@giacecco/the-data-ownership-delusion-4012cc232a2a

Similarly, the European Commission’s note on “What is personal data?” states that

personal data that has been de-identified, encrypted or pseudonymised but can be used to re-identify a person remains personal data and falls within the scope of the law. Personal data that has been rendered anonymous in such a way that the individual is not or no longer identifiable is no longer considered personal data. For data to be truly anonymised, the anonymisation must be irreversible.13

Concerns about appropriate use arise to the extent that these data are personal in the sense captured by the GDPR, on the one hand, but can be collected and arranged in patterns whose features can be exploited only through the use of sophisticated software and hardware, on the other hand. The raw materials that are ultimately being monetized or otherwise put to use originate in actions of individuals, who often do not perform actions with the intention of generating, or even with an awareness of their actions generating, input into patterns that would play a role on any kind of market. At the same time, it is only in virtue of technologically sophisticated recordings and transmissions that, somewhere else in the world, these data come together in ways that allow for any kind of exploitation. Electronically recorded behaviors make up a substantial part of many people’s daily activities. To the extent that these activities are recorded for the sake of commercialization it is their personal realities that are commodified. So, how can data be owned?

Absent any kind of legislation that makes different provisions, the data, once collected, would be under control of the entity that has gathered them – normally companies that provide smart phones, tablets, personal computers, digital assistants, electronically linked household appliances, positioning systems or search engines. Absent regulation to the contrary, data collectors would articulate some version of “finders, keepers.” The argument in favor of “finders, keepers” is that it is only in virtue of the computer-mediated character of many interactions that we can observe behavior that was previously unobservable, which enables transactions that were unfeasible before. The kind of regulation that contradicts such an approach would either constrain ways in which data can be obtained or aggregated to begin with or constrain the use to which data could be put once collected, or both. Either way, it is at this stage that the question of data ownership becomes significant.

Humanity has never had much difficulty in increasing the range of things that can be owned, and this would especially include things essential to creating wealth and status. Land, hardware, labor, ideas – all these have fit that description over the centuries. Each of them is regulated by a comprehensive set of laws accompanied by much philosophical reflection on the adequacy of any given set of such laws compared to alternatives, as well as of course much political mobilization around such alternatives. Data as they are understood in the digital world are newcomers in the domain of things that can be owned. Given the importance of data ownership and the novelty and elusiveness of the topic it is unsurprising that several answers were formulated in ways that analogized data to things with regard to whose regulation ownership considerations have long been prominent.


Let us consider the leading data-as proposals one by one. One attempt to bring data under the purview of ownership is to analogize them to natural resources. Obviously, data are not themselves such resources since they are generated through digital activities of humans. But data might nonetheless be relevantly like natural resources absent any independently given claims to them. Much like, say, oil, so this argument goes, data are just there. If we ask who should get to use them, the answer comes in terms of methods of extraction. It is those who do the work of collecting data who get to exploit them, much as it is those who extract oil who get to put it to commercial use. Since oil is the most referenced resource in this context, this proposal is called Data as Oil.13

For the last 150 years or so, oil has been crucial to enabling global transportation systems (in the form of gasoline, diesel, jet-fuel, home-heating fuel, lubrication oil or asphalt), not to mention the petrochemicals that enter into the production of plastic, synthetic fiber, drugs, soap or paint. Unsurprisingly, this train of thought has seemed attractive especially to industry representatives. After all, the analogy draws attention to the fact that, much as oil requires extraction and distillation to be useful, so data require collecting devices and mechanisms as well as mathematical analysis to be useful. The term “data mining” that is used for such analysis seems straightforwardly inspired by the analogy to oil.

However, the disanalogies between oil and data are glaring. Oil is relatively scarce, fungible (controlling for grade, oil from different regions is largely interchangeable) and rivalrous (can only be used by one party). Data is neither scarce nor fungible nor rivalrous. Most importantly for present purposes, Data as Oil runs afoul of the basic insight from the preceding section, that data are generated by human activities. In fact, by side-stepping the ways in which data are produced, this approach forecloses the sheer possibility of diagnosing any type of misappropriation. We


13. My discussion follows Scholz, “Big Data Is Not Big Oil.” Scholz also documents who has supported Data as Oil. According to Scholz, data scientist Clive Hunby coined the phrase “Big Data is the New Oil” in 2006.
should agree with Scholz in concluding that the only meaningful similarity between the data and oil is that they are immensely valuable for commerce.

Another attempt to bring data under the purview of ownership is Data as Labor. The motivation is that data should be owned by those who provide them: acts of generating and thus providing data are a type of labor. If so, such labor should be compensated in the variety of ways labor tends to be compensated in other contexts. In the first instance this would mean wages of sorts would be paid, one way or another, by companies that collect data from individuals. Such companies must find some ways of remunerating people for time needed to generate the data. To be sure, actual wages may not be practical, but one could think of other forms of compensation, such as privileges of sorts on the platforms though which data are collected that would be equivalents to wages. Moreover, if data is labor, then presumably the laborer would have control over whom to sell them to, and in what manner. Conceivably she would even have claims to a share in the profits made on the basis of her data contributions.¹⁴

The notion of labor that enters into this proposal is peculiar, especially when compared to the Marxian view of labor. On the Marxian understanding labor allows workers to leave their mark on the world. What Marx had in mind, however, was an actual and direct shaping with hands and minds, rather than something that comes about as a by-product. But more importantly, one problem is that the payment would be rather minuscule. What this implies in particular is that these payments would then not stand in much of a relation to what people do with the data and thus to the value of the product of the labor. For that reason, Data as Labor is implausible.

A third approach – Data as Personhood – insists that, in virtue of having been produced by humans, data express aspects of personhood, one way or another.¹⁵ For that reason, any kind of regulation that applies to them should be concerned with protection of personhood. Whereas Data as Oil and Data as Labor draw attention to particular ways data could in fact be owned (parallel to how natural resources could, or to how workers would be remunerated for labor), the most straightforward understanding of this third proposal encourages us not to think of data as being owned at all. Instead, their use should be guided by rights protecting personal integrity. Data generated by human activity extend personhood and ought to receive protection accordingly. There is also a different understanding of this argument, in terms of self-ownership, that would allow us to see Data as Personhood as a proposal about ownership.¹⁶

To be sure, property and personality are connected. (Hegel was basically right here.¹⁷) But this whole approach should be considered more as complementing ownership considerations pertaining to data, rather than as informing those very considerations. What brings data under the purview of ownership is that they have market value, and that value comes from collective activity. But there is then also this other (complementary) way of thinking about data: individually they are expressions of personality. These are potentially competing but also supplementary perspectives. But the Data as Personhood approach as such does not help with sorting out the ownership considerations.¹⁸

Next let us discuss Data as Salvage.¹⁹ Black’s Law Dictionary defines salvage as “rescue of imperiled property.” A salvage award is compensation for people who have helped rescue property, in particular property lost at sea. Those who salvage resources do have some claim to their value in virtue of the work they invested, without which the property likely would have perished. And to be sure, there is nothing wrong with what they did. Nonetheless, they are not normally the owners of the resources they have salvaged. As Scholz notes, this approach addresses the intuition that “data miners” should be compensated for work done to generate marketable outputs while also recognizing that the data themselves are traceable to somebody else. That is, without the work done by the “data miners,” these data likely would not survive as data. And even if they had, they would be of no value, and thus be lost to the commercial world. But prior to such work being done on them, they were recognizably somebody else’s.

This approach too runs afoul of the basic insight identified above, that what makes data valuable is that they are collectively generated. This approach focuses on how somebody other than the producer or original owner will have an ownership claim to something. But this approach makes no room for articulating the nature of the original thing that is owned and for illuminating how it becomes valuable.

Finally, let us consider Data as Intellectual Property. Typically, intellectual property are things such as scientific, musical, literary, or other artistic works and inventions, but also images, names, symbols, or design patterns. Recall that, by way of distinguishing data from oil, we noted that (unlike oil) data are non-rivalrous, non-fungible and non-scarce. These are features data share with intellectual property. In addition, both data and intellectual property are key assets in the knowledge economy. So to that extent, it seems plausible to treat data either straightforwardly as intellectual property or else at least in rather similar ways, a thought that has been taken up in European Law.²⁰ But there are

---

¹⁴ For Data as Labor, see Posner and Weyl, Radical Markets, chapter 5; Arrieta-Ibarra et al., “Should We Treat Data as Labor? Moving beyond ’Free.’” See also https://www.weforum.org/agenda/2017/09/the-value-of-data/

¹⁵ See e.g., Balkin, “Free Speech in the Algorithmic Society.”

¹⁶ This would take us to Nozick, Anarchy, State, and Utopia.

¹⁷ Hegel, Philosophy of Right. For discussion, see Waldron, The Right to Private Property, chapter 10.

¹⁸ For a critical take on Data as Personhood, see also Cohen, “Examined Lives.”

¹⁹ The discussion of Data as Salvage again draws on Scholz, “Big Data Is Not Big Oil.”

²⁰ Grosheide, “Database Protection—The European Way.”
important differences between data and intellectual property. To begin with, ideas are of value individually and separately, data only in large quantities. Moreover, legal protection of ideas is normally connected to some kind of creativity, a creativity beyond the ways in which algorithmic collection of data would be creative.

But these differences notwithstanding, I agree that, in all ways that matter for legal protection of ownership claims, data patterns just are like the typical kinds of intellectual property, and will proceed under that assumption. My main point going forward is that intellectual property as it is commonly understood (as I explain below, in the Lockean tradition) is over-protected, and that the reasoning behind that conclusion also speaks to data. This reasoning renders certain considerations applicable that I am about to introduce and that should make us reconsider the way intellectual property is handled in the first place.

In any event, what all these data-as proposals share is that they do not explain or even highlight what is valuable about data. This result leads us to my own proposal: Data as Collectively Generated Patterns. The idea is that the value of data does not consist in individual items but in the emerging patterns – and then not just for the people who provide them, but for lots of others relevantly like them. The value lies in the collectively generated pattern. Or anyway, this will be so in a broad range of cases. If in some cases in which data are valuable it is for some other reason, then what I argue here will still hold for that particular range.

By a collectively generated pattern, I mean this: “Under social parameters \( p_n, \ldots, p, \) as they prevail in country \( C, \) individuals with features \( F_n, \ldots, F, \) will with probability \( p \) do action \( A \) under circumstances \( C_n, \ldots, C, \).” This could be anything from individuals in a group determined by certain features hailing a cab or ordering pizza to marriage or home-buying behavior. I submit that data generated from the behavior of many individuals are valuable because, or to the extent that, they allow for a statistical inference towards that type of conclusion. Unlike the alternatives, Data as Collectively Generated Patterns does not create any equivalence with another domain where ownership is already well-understood. This approach makes clear how ownership considerations enter, but we must explore separately just how they do. I now propose a way for ownership considerations to bear on data once we understand them that way.

5. Intellectual Property: Locke vs. Grotius

To understand how ownership considerations enter my proposed way of thinking about what makes data valuable, for now we need to take a step back and talk about a seemingly very different topic: humanity’s collective ownership of the earth. From there we will then make a connection to intellectual property, and then to collectively generated patterns.

That humanity collectively owns the earth – and thus that the earth is a kind of Global Commons – was a predominant idea in 17th century political philosophy. Hugo Grotius, Samuel Pufendorf, Locke, and others debated how to capture this status and the conditions under which parts of that Global Commons can be privatized. These ideas mattered in the 17th century when the object whose appropriation mattered most was land and a crucial question was how competing European powers could legitimately appropriate land far away from home.

Locke famously merges his account of collective ownership with a labor-based (“mixing”) approach to privatization. Individuals could privatize parts of the divine gift of the earth by working on the land. They would mix their labor with the land, and thereby staking out a better claim to that land than any others would have. To be sure, there were constraints to how much of such appropriation there could be. “Enough and as good” had to remain in common possession for later arrivals to privatize land in the same manner; and people could appropriate only a quantity of land such that no spoilage would occur.

Subsequently, many commentators thought these ideas transferred readily to intellectual property. That is, parallel to how there is a Global Commons (and surface of the earth), so the set of ideas forms an Intellectual Commons. Those who “have” ideas are then not inventors or creators, but explorers or discoverers. What claims to controlling use of ideas there can be would have to be evaluated in light of the fact that such ideas originally belong to a Commons. But then the ideas about “mixing labor” that originated in the debate about land ownership could be re-articulated to allow for a far-reaching right to privatization in the domain of ideas.

The mixing itself would consist in whatever labor was necessary to develop an idea into some kind of marketable product. The constraints – to leave “enough and as good,” and to not generate spoilage – are satisfied straightforwardly in the case of an Intellectual Common. After all, there plausibly would be infinitely many ideas, so there will always be “enough and as good left” for any finite number of humans who might want to exploit the Intellectual Common. And unlike land there is no sense in which intellectual products could “spoil.” Locke’s discussion in Chapter V of the Second Treatise of Government has been said to have “totemic status” in theorizing intellectual property.

However, this transfer of ideas from the domain of land acquisition to intellectual property is rather regrettable. Locke did not take much of an interest (if any) in the possibility that that there was good reason not to accept privatization of certain parts of the collectively owned earth, especially the seas. As opposed to that, it is for his work on the morality of appropriating the sea, and then specifically his arguments against such a possibly, that Grotius is best known. Throughout his writings, he argues in different ways that the seas cannot be owned. The sea is free in the sense that all human uses of the sea, ranging from fishery to mere passage, were permissible for everybody who could get there. And I submit that it is precisely to those parts of the surface of the earth that the realm of ideas is analogous.

Today, complete freedom of the seas as he envisaged it in the 17th century would no longer be called for on Grotius’ own terms. Yet Grotius’ reasoning bears on a rather different domain, the products of the mind, such as scientific, musical, literary, or other artistic works and inventions, but also images, names, symbols, or design

21. See Buckle, Natural Law and the Theory of Property; Tuck, The Rights of War and Peace.
22. Locke, Second Treatise of Government. For discussion, see Waldron, The Right to Private Property, chapter 6; Sreenivasan, The Limits of Lockean Rights in Property.
patterns. These products are subject to intellectual property law, which, among other things, includes patents, copyrights, and trademarks. To be sure, a Grotian approach to intellectual property law would be consistent with there being compensation for those who transform ideas into marketable products and with setting incentives for such work. But the main point is that, once we think of ideas as parallel to water and of water as not amenable to the kind of activity that would allow for privatization, the same would be true of ideas as well. Ex ante they belong to everybody, and it is hard to see how there could be benefits to developers of ideas beyond what I just sketched. A Grotian approach delivers much more restrictive private intellectual property rights than a Lockeian approach.  

So let us explore how Grotius’ arguments against private appropriation of the sea carry over to this case. To begin with, Grotius points out that use of the sea is consistent with everybody else’s use of it. That would be true of ideas as well, and in fact of ideas more clearly than of water. Thomas Jefferson classically makes that point about intellectual property in a letter from 1813:

> If nature has made any one thing less susceptible than others of exclusive property, it is the action of the thinking power called an idea…. Its peculiar character... is that no one possesses it the less, because every other possesses the whole of it. That ideas should be freely spread from one to another over the globe, for the moral and mutual instruction of man, and improvement of his condition, seems to have been ... designed by nature... Society may give an exclusive right to the profits arising from them, as an encouragement... to pursue ideas which may produce utility, but this may or may not be done, according to the will and convenience of the society, without claim or complaints by anybody.  

There is a point to having private property in things like apples since only one person can make certain kinds of use of them. As Grotius insists with regard to the sea and Jefferson with regard to ideas, however, there is no such point in having private property rights in either of these. Crucially, gains for occupiers, certainly in the case of ideas, do not depend on excluding others, if we talk about the actual use of ideas, rather than profits accrued from the exclusion.

Secondly, *Mare Liberum* argues for freedom of the seas by appeal to its relevance for trade to establish that *everybody benefits* from leaving the seas free:

> For even that ocean wherewith God hath compassed the earth is navigable on every side round about, and the settled or extraordinary blasts of wind, not always blowing from the same quarter, and sometimes from every quarter, do they not sufficiently signify that nature hath granted a passage from all nations unto all?  

Similarly, use of ideas by some subtract nothing from their usefulness for others. On the contrary, it adds to it, by increasing the overall amount of intellectual activities that in turn would inspire yet more such activities and thereby also increase the availability of whatever benefits such activities may have. *Everybody* benefits from a situation in which ideas are left unappropriated (given, also, that anybody’s use of them does not interfere with everybody else’s use), whereas only a few will benefit, respectively, if appropriation of ideas is protected by social and legal norms.

Let us proceed to the third point Grotius makes about the seas: the seas cannot be meaningfully occupied. One cannot do anything to water parallel to how, in the case of land, “the beginning of Possession is joining Body to Body” (*DJB*, II.8.VI). A body A’s being joined to a body B decreases the space for a body C to be joined to B. Such joining might either affect the object itself in a way that would make it impossible for others to join the object in the same way, or else it would create a situation where others could join the object in the same way only by violating more basic moral rights of the person who did the original joining (say, because she needs to be pushed away). These considerations do not apply to water, and thus water cannot be meaningfully occupied.

In straightforward ways of understanding what it is to occupy something, it is in a parallel manner true of ideas that they cannot be occupied. To be sure, one can keep ideas secret, or distract people from them, but one cannot do anything to an idea that keeps it from being independently grasped by others. And a mind’s grasping an idea decreases no other mind’s capacity to do the same. A mind’s grasping an idea does not affect the idea itself in ways that would make it impossible for others to grasp that same idea, nor does a mind’s grasping an idea create a situation where others could do the same only if they violated more basic moral rights of the original grasper. Like water ideas cannot be meaningfully occupied.

---


27. Of course, were we to change intellectual property arrangement now, some would indeed be made worse off by such changes (so not everybody would benefit from these changes), namely, those who so far had been allowed to appropriate ideas, respectively. Yet what I have argued does hold from an *ex-ante* standpoint in which no intellectual property arrangements have been made yet and from which we must assess what sort of private rights to intellectual property (if any) there should be.


29. One might object that one can indeed “occupy” ideas in the sense that there could be (and in fact are) norms of intellectual ownership, such as patent law and copyright law. Alas, “occupation” of ideas is possible only through the acceptance of such norms -- norms that require of all people other than the holder of the respective private property right to renounce the option of making use of ideas although their making such use of ideas could occur consistently and simultaneously with everybody else's making that same use of ideas. This observation then raises the question of why anybody ought to accept such norms, a question that, in turn, takes us back to the other two considerations against privatization we already discussed.
6. The Intellectual Commons

What we have seen so far is that, for intellectual property, certain considerations would support limitations on privatization were there a presumption against privatization. A ready way of arguing for such a presumption is to show that there is an Intellectual Common in the same way in which there is a Global Common. A straightforward way of arguing that, in turn, is to defend a kind of realism about intellectual products.

Such realism denies that scientific, musical, literary, or other artistic works are literally “products” of the mind. Instead, they exist outside the realm of either material or mental objects. They belong to a (Fregean or Popperian) “third realm” of non-mental super-sensible entities, distinct from both the sensible external world and the internal world of consciousness. Alleged “products” of the mind would be such products only in the sense that a conscious mind can discover them. There would be no invention, no refinement, or any other contribution to these entities. This view delivers a presumption against privatizing elements of this third realm. For objects in that realm exist prior to any human activities. In a second step we could add the considerations against privatization we extracted from Grotius’ discussion of the sea, to show that this presumption is hard to overcome.

To be sure, this presumption can be overcome. First, individuals may fairly claim compensation for investments in making ideas accessible, compensation that might take into account the particular opportunity costs of the relevant individuals. Second, consistent with this argument for limitations on private intellectual property rights is for societies to set incentives to stimulate creativity. Acknowledging compensation and incentive-setting as reasons for creating private intellectual property rights, we still leave open much potential for disagreement about how far-reaching these considerations create – a point we will not further pursue here but need to acknowledge.

So far, we have assumed that there is an Intellectual Commons much like a Global Commons. By way of contrast, consider a characterization of intellectual products that overemphasizes the subjective aspect mirroring how our earlier characterization overemphasized the objective aspect. According to this characterization, intellectual products are not discovered, but invented and created. There is no Fregean or Popperian third realm, no Intellectual Common, no presumption against privatization.

Once we think of ideas as parallel to water, and of water as not amenable to the kind of activity that would allow for privatization, the same would be true of ideas as well.

We cannot even state that we instead have a presumption in favor of privatization because there is no starting point with regard to which anything could be privatized. But now it looks as though we have a presumption in favor of private property rights, potentially much beyond what compensation and incentive-setting license.

But crucially, and perhaps surprisingly, what we above identified as the three Groton considerations against privatization reenter. These considerations were: that ideas cannot actually be occupied in the same sense in which, say, land can be occupied; that the gains for users of ideas do not depend on excluding others; and that leaving ideas unappropriated benefits everybody. Above, these considerations ensured that the presumption against privatization could generally not be overcome. (The exceptions were fairness-based compensation and consequentialist considerations in favor of incentives for invention.) Now the considerations against privatization reenter by limiting the extent of rights for which in this case there is a presumption. These considerations again ensure that we consider the standpoint of those expected to comply with intellectual property law. Both above and here these considerations entail that we should limit private property rights to what we can obtain via appeals to fairness and incentive-setting, although they enter in rather different ways.

So we have operated with two caricature views on the ontology of the objects of intellectual property. The realist account unduly eliminates the contribution of human creativity, whereas the anti-realist account overstates the role of individual minds. But as we have seen, the same results follow for intellectual property regulation regardless of whether we have a third realm of ideas or whether ideas are human creations. So we can state the main result of this discussion of the possibility of private intellectual property rights as follows: The ontological status of particular intellectual products will have to be characterized to some extent in terms of components readily placed into a third realm.

30. Gottlob Frege’s 1918 essay “Der Gedanke: Eine Logische Untersuchung” (“The Thought: A Logical Investigation”) is a locus classicus for this view – Frege, “Der Gedanke. Eine Logische Untersuchung.” Although I am, for the sake of the argument, offering an extreme version of it. See Gideon Rosen’s entry on abstract objects at the online Stanford Encyclopedia of Philosophy, http://plato.stanford.edu/entries/abstract-objects/. For a general background discussion of abstract objects and questions of their existence, see Burgess and Rosen, A Subject With No Object. Important to mention is also Karl Popper’s theory of reality, which distinguishes among three worlds: World 1 is the world of physical objects and events; World 2 the world of mental objects and events; and World 3 is the world of the products of the human mind. See for instance Popper, Objective Knowledge.

31. “We may begin thinking about information rights, as Hettinger does, by treating all ideas as part of a common pool and then deciding whether and how to allocate to individuals rights to items in the pool. Within this framework, ideas are conceived on the model of tangible property. Just as, in the absence of social institutions, we enter the world with no particular relationships to its tangible assets or natural resources, we have no particular claim on the world’ ideas. In this scheme, as Hettinger asserts, the burden of justification is very much on those who would restrict the maximal use of intellectual objects.” (p 20) Alternatively, we may begin, as I do, by thinking of ideas in relation to their originators, who may or may not share their ideas with specific others or contribute them to the common pool. This approach treats ideas as central to personality, and the social worlds individuals construct of themselves. Ideas are not, in the first instance, freely available natural resources. They originate with people, and it is the connections among people, their ideas, and their relationships with others that provide a baseline for discussing rights in ideas. Within this conception, the burden of justification is on those who would argue for disclosure obligations and general access to ideas;” Paine, “Trade Secrets and the Justification of Intellectual Property: A Comment on Hettinger,” 49. This is a response to Hettinger, “Justifying Intellectual Property.”
and to some extent by appeal to human creativity. (One of these extents may be vanishing.) So to the extent that we must appeal to something in that third realm, the considerations used for that case apply; to the extent that we are talking about products of the human mind, the considerations given in that case apply. Either way the respective argument generates the same constraints on private rights. Therefore, these constraints apply to the whole range of intellectual property.  

7. Ownership of Collectively Generated Patterns

The Grotian considerations formulate general constraints on appropriation in any domain of ownership. They no longer fit the seas well. But for ideas, I submit, these arguments succeed, and deliver a more plausible theory than the dominant Lockean approach. Let us reconnect now to our discussion of Data as Collectively Generated Patterns.

Recall that data-as proposals aim to do three things. First of all, they make a suggestion for what it is about data that makes them valuable; secondly, by doing so the proposal points to who should own the data; and thirdly, the proposal also points to how data should be owned, to the kind of rights involved. Data as Collectively Generated Patterns straightforwardly does the first. For the main result from section 6 to tell us then how things can be owned, we must show that that result applies to collectively generated patterns. And finally, we must explore what can be said about the owners of collectively generated patterns.

What we said above was that the ontological status of particular intellectual products will have to be characterized to some extent in terms of components readily placed into a third realm, and to some extent by appeal to human creativity. That discussion proceeded in terms of the ontological status of intellectual products, and the extreme views that allowed us to frame that discussion were, first, that ideas (and other intellectual products) exist in ways that are entirely mind-independent, and secondly, that ideas (and other intellectual products) come into existence in the mind of a creator or inventor. Either way, the same constraints on private rights are forthcoming. Intellectual products will generally be some kind of mix of the considerations that these extreme views emphasize, and so regardless of their precise ontological status these same constraints on private rights apply.

Above we introduced the following formulation to capture what is meant by collectively generated patterns: “Under social parameters \( P_1, ..., P_n \) as they prevail in country \( C \), individuals with features \( F_1, ..., F_m \) will with probability \( p \) do action \( A \) under circumstances \( C_1, ..., C_n \).” Such patterns are the kind of objectively given entities to which there should only be limited property rights anyway. In other words, these patterns either are, or are sufficiently like, entities in a Fregian and Popperian third realm to be straightforwardly governed by the Grotian considerations. But suppose somebody objected that this way of characterizing collectively generated patterns neglected the subjective contributions of individual actors to their creation. To be sure, in this case the subjective part would actually consist of two components: the individual actions that partially constitute the patterns, on the one hand, and the activities of the firm to make the patterns visible, on the other. In other words, the part of the inventor in the general domain of intellectual property here is filled in two kinds of ways. In response to such an objection then the “subjective wing” of our main result enters, as a fallback. Either way, our main result would capture everything one needs to say about this case. That is, regardless of the precise ontological status of collectively generated patterns these same constraints on private rights apply. In this way we can see that both the users and the firms are constrained in terms of what they can do with ownership of data in much the same way in which the producers of ideas are constrained.

Unlike the other data-as proposals, mine is actually not primarily an ontological approach that highlights the nature of the data. Instead, the focus is on characterizing the ways in which data fit into human practices of assigning commercial value to entities. So the proposals are more pragmatic than ontological in nature. In fact, the very use of the term “data” (what is given) indicates a lack of ontological questions by way of contrast with the pragmatic aspects of what is done with what is given. Nonetheless we have ended up with ontological investigations after all. And it is partly precisely because my proposal is not ontological in nature the way other data-as proposals are that we needed to do this in-depth investigation of how ownership considerations enter—which then after all involved us in ontological investigations.

Nonetheless, the view developed here does not presuppose any particular view on the ontology of data since, per the main result from the last section, the Grotian considerations apply regardless of the objective or subjective nature of the collectively generated patterns. Accordingly, private property rights to collectively generated patterns are limited to compensation for those who transform such patterns into marketable products and with setting incentives for such work. Much like in the more general case of intellectual property, corporations should claim proceeds from innovation to the extent necessary to compensate them for their effort and not to stifle innovation. But that would be the extent of it.

So now that we see what kind of private ownership rights are forthcoming for collectively generated patterns one remaining question is precisely what collective is it that should own? In principle there are various candidates here: the people involved in the behavior captured by the regularities; all of the people in a certain country that generate the context in which these regularities become possible in the first place; or everybody in the world. We can safely exclude the first option according to which it would just be the people immediately involved in the generation of the pattern. After all, they are doing what they are doing only

32. To use a mathematical analogy: We have offered an argument for two extreme cases, and now have argued that the same argument also holds for the intermediate cases that can be understood as convex combinations of the extreme cases. X Perhaps see here for additional references – https://plato.stanford.edu/entries/intellectual-property/

33. Nowadays there is much overuse of the seas, especially overfishing. The idea that use by one person is consistent with that of everybody else is limited to a time when technology could not do much to the sea. Moreover, in contrast to the 17th century, in the present era there are ways of appropriating the seas, with appropriately involved use of technology. So that leaves us with the quasi-utilitarian consideration that a free sea enhances collective well-being.
before the background of parameters that prevail in their wider context. So the relevant debate would be that between everybody in a country and everybody in the world. And the best response to that question would be a general reference to the justifiability of countries.  

8. Conclusion

To be sure, our discussion here has not covered all types of data collection. There is a type of surveillance that is exclusively concerned with one person and is used for the sake of controlling that person rather than for the sake of making predictions, or in any event does not involve the behavior of other people and thus does not turn on patterns at all. This kind of data collection is not covered by my approach. My approach here neither is, nor is it intended to be, all-encompassing. But it does cover a lot of ground when it comes to regulation of data ownership.

Given how important data ownership will be in the future there is much at stake when it comes to sorting out who gets to own data. The current default is that data, once collected, are owned by the entity that has gathered them – normally companies that provide smart phones, tablets, personal computers, digital assistants, electronically linked household appliances, positioning systems or search engines. My argument in this paper has been that this is a highly unsatisfactory default that we should change: the default should be that collectively generated patterns should be collectively owned, in ways that would then allow for individual claims, liberties, powers and protections to be sorted out in a next step.

That is, those who provide the means to collect the data that reveal such patterns should be fairly compensated for their services, and there should also be financial incentives of sorts for companies and individuals to do the kind of work that makes such data collection possible. But what would be ruled out is that the companies are involved with data collection have unlimited and exclusive control over them that would allow them to anticipate societal trends or even redirect behavior in ways that other actors could not do with those same data. At the same time, those whose data are being collected are then also entitled to consideration, especially to the kind of consideration captured under “privacy.”

For illustration, recall the three cases from above. We can be sketchy in all cases since the point is just to indicate how one could think about such cases now that we see the data are collectively owned. In Case 1 data analysis reveals that a certain pattern of typing on keyboards indicates the onset of a debilitating muscle disease. Insurance companies pay for such information because it allows them to reject potential clients or adjust premiums. In this case a context-specific understanding of privacy (following Helen Nissenbaum) should be operative and lead to a prohibition of this kind of use of data.

In Case 2, data analysis reveals that someone’s reading tastes make it likely that she would be receptive to advertising for a new book on internet privacy. Internet sellers that carry that title pay for that information to improve marketing of their product. In this case, because of the trivial nature of the case, no regulation would be needed but the data should be broadly available so that others could also advertise this way.

In Case 3 data analysis renders it likely that somebody is the kind of voter who could be persuaded to favor a particular candidate in response to a certain type of input (regardless of whether that input is accurate, which she would not be able to verify). Political campaigns pay for such information because it increases their chances at winning elections. Here it would be in the interest of democratic politics if such data were spread broadly. And, of course, the accuracy of the input would need to be discussed under a different heading.

So to be clear, the fact that collectively generated patterns are collectively owned must not mean that everything is made available to everybody indiscriminately. So neither is it the case that companies that offer services to collect data remain without any payoff at all, nor is it the case that individuals would not be entitled to moral consideration in the form of claims, liberties, powers and protection. But all these matters are to be sorted out before the background of collective ownership in these patterns. And to make something like this a reality we may need a very different internet.

34. And this could be done, for instance, in the style of an appeal to the limitations of utopian reasoning, see Risse, On Global Justice, Part IV.

35. Nissenbaum, Privacy in Context.
Literature


