IARIW – CIGI 2023

IARIW – CIGI 2023

Thursday, November 2 – Friday, November 3

Data as Representation

Sean Martin McDonald
(Centre for International Governance Innovation)

Ben Gansky
(Arizona State University)

Paper prepared for the Conference on The Valuation of Data November 2 – November 3, 2023
Session 6: The Valuation of Personal and Health Data
Time: Friday, November 3, 2023 [10:00 AM-11:30 AM EST]

Data as Representation

By Sean Martin McDonald and Ben Gansky

Introduction

There are a lot of ways to try to understand the value of data, economic and otherwise, and they vary substantially in both method and purpose (see, e.g., Beauvisage & Mellet, 2020; Birch et al., 2021; Collington, 2019; Parsons & Viljoen, 2023; Vertesi & Dourish, 2011). Our analysis does not attempt to develop a competing comprehensive theory of data valuation or typologies, nor do we describe a method for the derivation of the value of data as commodities or assets, nor does our analysis aspire to promote any particular market or regulatory structure. Instead, this paper focuses on the most common use of data - as an assertion of a fact in a context designed to make or influence a decision - to reframe what it might mean to identify the value of data. This specific kind of data use to which we refer is equivalent to a legal term of art: representation. When people share and use data they are, more often than not, making a representation, meaning an assertion on the basis of which a decision is to be made. By drawing on this legally defined concept of representation, focusing especially on rights-impacting contexts, we contextualize the conversation about data value in terms of what makes assertions valuable, under what circumstances, and for whom.

Data are not fungible

While data also has other uses and contexts, the fundamental difference between data and other units of exchange is that data is information. This basic fact has been in recent years somewhat confused by the popularization of discursive framings of data (e.g., as oil, sand, or plutonium; cf. Doctorow, 2008; O'Reilly, 2021) which metaphorize data as a fungible market commodity. The idea that there could be such a thing as a standard unit of data is partly responsible for engendering a wave of failed consumer-to-business data projects (Beauvisage & Mellet, 2020) – to say nothing of the misapprehension of data as 'personal' rather than social (for a historical account of this discourse, see Igo, 2018; for a rigorous theoretical account, see Viljoen, 2021; Parsons & Viljoen, 2023). Critiques of this kind of data-as-commodity thinking have more closely examined the actual practices of data-entangled corporations (Birch et al., 2021) and noncommercial projects alike (Vertesi & Dourish, 2011) to demonstrate that data are in fact highly differentiated and their exchanges conditioned by cultural and infrastructural, as well as economic, factors. Nevertheless, data valuation discourse has produced its own material effects, notably by making data (illusorily) appear as a familiar sort of object within a capitalist schema, whether commodity or asset – triggering a knee-jerk response towards accumulation (cf. Sadowski, 2019).

This recent history has tended to obscure the central importance of the unique relationship between value and information for efforts to establish the value of data. It should come as no surprise that some of the most clear-eyed accounts of this political economic relation are situated in studies of data in advertising and marketing – economic sectors which found a business model for data exchange in the 1970s (i.e. targeted advertising) and have ridden that arrow up and to the right in the decades since (Hwang, 2020; McGuigan, 2019; McKelvey, 2022; Turow, 2012, 2017; Zuboff, 2015). The informational content of data about advertising targets is of value to advertisers in so far as it affords them a higher degree of confidence that their intended targets will be receptive to their messaging, and provides feedback about targets' behavior after being exposed to these messages. How much value advertisers place on this information and the ways in which targeted advertising platforms make use of it depends on the confidence advertisers have in the validity of platforms' assertions – that is, that platforms are giving true and useful representations. If this confidence were to plummet, so too would the perceived value (and actual price) of ad placement inventory (see Hwang, 2020 for a detailed assessment of this dynamic and its precarity).

When we write that data, as a vehicle for information, is non-fungible, we mean that there is no standardized unit of data that we can treat as interchangeable with other units. Information is valued contextually – this is demonstrated by the ways in which data brokers are able to profitably arbitrage transactions across significantly different data contexts, flexibly collecting and exchanging between contexts where data collected, repackaged, and sold including, e.g. granular location data, commercial and financial transaction data, census data, online behavioral data, and law enforcement data. The contextual valuation of data is also demonstrated by the few areas of collective life in which this flexibility is curtailed: for instance, in the United States, certain forms of health data and information about students are protected from certain forms of collection and exchange. In these contexts, what makes information valuable is conditioned, inter alia, by what regulators will permit; if data is illegal to collect, package, and use in particular ways, those restrictions will condition the value of that data in that context (for instance, by either making it worthless in other contexts, or much more valuable precisely because it is illegitimately come by). There is neither a unified market for 'all data' nor discrete, cleanly segmented exchanges. Straightforward economic approaches to assessing data by using price as a proxy for value are bound for frustration.

Contextual factors conditioning the value of data as assertions of fact

The more sensitive the context, the more important it is for decision makers to have confidence in the data they use as representations. The salient aspects of context here are crucially not only about the *situation* in which data are mobilized to make decisions and the *impact* those decisions might have on affected individuals and communities. Our analysis draws particular attention to the nature of the *relationships* between data supply chains, individuals and social constellations affected by data-derived decisions, and the representatives responsible for decision-making on behalf of these affected subjects.

All data are not equal; what makes some data more valuable than others in a particular context is the perceived quality of the supply chain by which the data are produced and through which they travel. The 'quality' (and therefore, value) of a data supply chain cannot be assessed acontextually. There are different standards for the representation made by a waiter who offers advice about the daily lunch special compared to the representation made by a pathologist about the best course of disease treatment – and we should expect a correspondingly different assessment of what makes for a more or less valued assertion in each context. Understanding the value of data use as a representation not only requires examining the 'technical' quality of data and its claims (towards useability and veracity), but also political contexts of production, the interests of those involved in the supply chain, and the substantive legal liabilities that result. In other words, the value of data at any stage of use is directly related to the legitimacy, stability, and quality of the underlying relationships between the data subjects, representatives, and the context of use that comprise the data supply chain. As many have noted, data use is not only importantly contextual - and contextually valuated - it is relational (Ferryman, 2017; McNealy, 2021; Viljoen, 2021). In other words, in addition to the technical and substantive questions that data-enabled representations raise, such representations also raise questions around whether the person or organization using data is the appropriate party to be making those assertions, in the context of use. But what exactly do we mean by a 'data supply chain'?

Data supply chains: their comparative epistemologies and politics

Data are never the product of immaculate conception; they are created, transformed, mobilized, and reconfigured at specific sites, each featuring particular actors, using particular instruments, for specific purposes (Baker & Millerand, 2007; D'Ignazio & Klein, 2020; Leonelli & Tempini, 2020; Pine & Liboiron, 2015). The specific qualities of the assertions made by a given dataset or data stream are conditioned by these supply chains. Few scholarly fields have made as sustained a research program of examining these dynamics as science studies. Over the past two decades, scholars have developed a body of evidence and theory converging on a number of shared propositions regarding the politics and epistemology of data practices and infrastructures.

Raw data is an oxymoron (G. C. Bowker, 2005; Gitelman, 2013). All data are partial inscriptions of selected phenomena; the captured phenomena and their selectively-represented aspects are those designated of interest in a particular context by a motivated actor with the resources to produce a representation in the form of data (Bogen & Woodward, 1988; Knorr-Cetina, 1999; Leonelli, 2015; Morgan, 2020; Ottinger, 2010). Even at the point of creation, data embed subjectivity by virtue of what they include and leave out (Almklov et al., 2014; Aronova et al., 2017; G. Bowker & Millerand, 2008; Lemov, 2015). The designation of fields and classification of items in a spreadsheet or form is anything but apolitical or objective (G. Bowker & Star, 1999). Data also embed theory; data are often purposefully 'normalized' according to a predetermined model of the phenomena in question (Bokulich, 2020; Bokulich & Parker, 2021; Leonelli, 2019).

This is often a necessary step in achieving the purposes for which the data is intended (Edwards, 2010). When data is contained within a single site for creation, transformation, and analysis, the quirks of its genesis and modifications may be understood and compensated for as tacit knowledge within a community of practice (Edwards et al., 2011). The truth-value of a data-based representation is contextually valuable; what counts as sufficient fidelity in one context will be different from another (Gressin, 2023). However, the entire premise of 'big data' as a novel form of knowledge production is predicated on the mobility and re-use of data beyond their context of origin (Bates et al., 2016; Borgman et al., 2019; Koch et al., n.d.; Tempini, 2020; Thylstrup et al., 2022). When data are then concatenated, munged, or otherwise combined, knowledge of their embedded limitations and subjectiveness is frequently irreversibly lost (Benjamin, 2019; G. C. Bowker, 2005; Braun, 2014; Chun & Barnett, 2021; Lauer, 2017).

For some, the slippage and opacity introduced by complex supply chains for data is useful. Hwang (2020) describes how providers of targeted advertising capitalize on the inscrutability of their reported performance metrics to preclude their customers from being able to closely examine the efficacy of their advertising purchases. Other purveyors of data products, such as data brokerages, strategically obscure their 'mix' of data sources as trade secrets (Crain, 2018; Katyal, 2019; Obar, 2020; Thylstrup et al., 2022). We have previously referred to this digital political economy as 'the supply chain shredder' (Gansky & McDonald, 2022). If we examine communities of practice for whom representational integrity is paramount, we observe different structures and practices of accountability and data maintenance.

The Intergovernmental Panel on Climate Change, for instance, is not only responsible for preparing reports for governments globally, but for managing the vast network of actors and infrastructures responsible for collecting, combining, and analyzing climate data in myriad forms (Edwards, 2010). Given the extreme political sensitivity of climate change science, the scientists, technicians, and managers involved have taken extraordinary measures to document and continuously manage the integrity of their data supply chains. The massive investment in these data supply chains yields returns not in the truth-value (or even the use-value) of any particular dataset they release, but in the continuous assurance of the integrity of the chain itself. This investment is contextually justified by the kinds of claims made by those who would rely on IPCC reports as valid assertions of fact: nothing short of the global renovation of energy economies. The general point is: the scope and impact of the decisions to be made based on data representations creates correspondingly weighty obligations to care for the supply chain which produces, shapes, and mobilizes that data.

Relationships as context for the value of data representations

As stated above, data's value and impact are not only related to the integrity of the relationship between the parties involved in the data supply chain, they are related to the context for that relationship. You may have totally appropriate relationships with your doctor and your lawyer, for example, but that doesn't mean it's appropriate for your doctor to represent you in court or your lawyer to make decisions about your medical care - even if the information they bring to each representation is factually accurate (see, foundationally Nissenbaum, 2009). The act of data sharing, and use, not only creates these relational and representational bonds between data users and producers as links in a digital supply chain - it also creates those relational and representational ties between data subjects and data users. On one hand we have the practices and structures assuring integrity and oversight – broadly speaking, situated accountability – for a data supply chain. On the other hand, we have the integrity of the relationships in the context in which data are employed as representations. The degree of fit between these two dynamics is what constructs the value of data.

Regardless of the intended impact, the exchange and use of data is impossible to extricate from the broader political economic and power relationships that define the relation between a service provider and beneficiary. The appropriateness of data exchange must be understood in the context of the rights that emanate from those relationships, regardless of whether those relationships are primarily or entirely digital. Historically, assessments of data - and the supply chains that underpin its aggregation and use - have focused on functional and qualitative characteristics, correlating data's assumptive value to its completeness, 'quality', and volume. As the public use of data to achieve governance outcomes has matured, both in theory and practice, it's increasingly clear that the value - and liabilities - created by data are increasingly defined by a different set of qualitative characteristics, like the diversity, legitimacy, and ethics of the underlying digital supply chain.

This shift from useability to politically normative rubrics implies a sensitivity to the rights and responsibilities emanating from the relationships encoded in data. Relationship models, as others have noted, are an integral component of the normative character of digitally intermediated relationships - most often articulated and enforced by an ecosystem of institutions. In particular, duty bearing professions - those that are regulated by public and private institutions - provide models for the way that we might govern digital representation relationships; they are realized by institutionally regulated, tangible, operational infrastructures designed to ensure the integrity, equity, and symmetry of power in inherently asymmetrical representation relationships (Balkin, 2020; Richards & Hartzog, 2015, 2020). While there is a significant range of practice, both within and between duty bearing professions, there are common governance design patterns that offer valuable guidance for those attempting to design integrity measures for data and digitally intermediated relationships.

This article proposes an approach focused in three parts: (1) the articulation and limitation of animating purpose - essentially, the rubric for determining the legitimacy of use in a context; (2) the relationships between the scope of representation, standards of expertise and care, and boundaries of available representative actions; and (3) the responsibility to support, if not provide, means of independent oversight and accountability. These broad dynamics won't address a commodifying approach to data governance - rather, they will provide ways to contextually assess the value and risks of the different approaches to building data supply chains, based on the highest-integrity models implemented in relevant contexts.

Data as a representation, by a relationship, in a context

This analysis starts from a few assumptions, worth articulating upfront - the first and most important of which is that the primary purpose of data use is as information in order to inform a decision. Those links can be direct - like when a witness gives testimony in a court trial; indirect - like when a person's medical test results go on to inform research beyond their care; and at a huge range of abstractions - like when a machine learning model is trained on massively aggregated data sets about population-level behavior. Even data products that don't target, limit or direct who their users are, like weather predictions, are designed with the recognition that people use them in order to inform decisions - whether what to wear or where to make industrial agricultural investments (e.g. Bates, 2014).

The primary variable in data's value is not whether the purpose of its use is to influence decisions, it is how explicitly designed and governed the underlying supply chain is - from data production to use. You cannot, for example, hold a weather prediction service liable for the ways you may have suffered from relying on their data - even if it was wrong because of a flaw in their system, whereas when a hospital uses digital diagnosis tools in order to determine a course of treatment for a patient, the accuracy of the outcome can create legal liability. The difference between the two is how specifically the data system is designed to influence the decision it's being used for and how directly connected the supply chain of stakeholders is to the data's ultimate use. Patients are able to hold their doctors directly accountable for their outcomes of their work, because governments and the medical profession recognize the importance and potential impacts of allowing bad tools and information into healthcare services.

Generally speaking, the more impact a decision has on the life of an individual, the more likely there are to be regulations, professional duties, and other practical protections focused on how those decisions are made - including what information can (and can't) influence those decisions. While deeply flawed, one example of this dynamic is evidence law - which, essentially, determines what kinds of expertise, what types of analysis, and what sources of information a judge or jury should consider when deciding cases (see, e.g. Rappert et al., 2022). Most legal systems recognize that the decisions made inside of courtrooms have large, often life-altering impacts, and that without strong rules to ensure the integrity of the information that shapes those decisions, politics and power can easily overcome justice. Evidence law includes in its rules and regulations that there be clear and extremely high-integrity linkages between the creation of evidence and its use in a courtroom, whether literal 'chain of custody' for physical evidence, a clearly stated base of expertise from witnesses, or specific limits on what experience entitles a person to make different claims (cf. Regulation of the European Parliament and of the Council on European Production and Preservation Orders for Electronic Evidence in Criminal Proceedings and for the Execution of Custodial Sentences Following Criminal Proceedings, n.d.). Evidence law has also, notably, struggled with evaluating the explosion of data and products changing the administration of law at nearly every level - from the use of new types of data, like biometrics, to new forms of analysis, like using mobile phone data to determine location, to new forms of fraud, like ensuring evidence isn't tampered with through deepfakes or security breaches (e.g. Hussain & Bowker, 2021; Lightbourne, 2017; Wexler, 2017). Evidence law not only regulates what data and information are allowed to impact decisions about peoples' freedom, they regulate who is able to be a source of that information and, critically, why.

Legal systems have a built-in recognition and mechanism for ensuring that information that shapes its work is appropriate for the impact of the decision it's used to make. Importantly - and obviously - evidence law doesn't apply until a lawyer attempts to use a piece of information in a courtroom. In other words, all of the rules that ensure the integrity of the information, data, and technologies that are allowed to influence legal process, are triggered by the context of their use, as opposed to their production. When evidence is introduced in court, it's explicitly done in order to make or contribute to a factual determination, with clearly stated interests, in direct recognition that they are likely to face contest, if not direct contradiction. And, so one of the critical ways that most legal systems protect people is to ensure that any information that can be used against you, comes from a source that can be articulated and confronted. Said a different way, the law's right to confront your accuser is an important foundation for understanding the quality of data supply chains - not because of the quality of the data they produce, but because of their fitness for use in high-impact, and thus, high-value contexts.

This section takes fitness for high-impact and high-value use as a framing assumption for our analysis, and uses existing models for high- integrity representation relationships to reverse engineer the core characteristics and operational requirements for valuing data supply chains.

Data as a representation

One of the definitional ambiguities embedded in starting an analysis from the term "data" is that it can strip away inherently important aspects of information exchange, like the source of that information or its intended use. Consequently, the term "data use," gets used as a euphemism for an incredibly broad range of information exchanges and behaviors, making it nearly impossible to make any kind of observation that applies universally. As will be obvious by now, our analysis starts from the "endpoint" of data use - the point at which data is used to influence a specific decision. By starting from the perspective of the decision data influences, instead of the influencing data, we are able to identify and reverse engineer a valuation framework for the supply chain of data inputs. This reframing enables us to avoid deriving contextual significance from the characteristics of given data and, instead, to focus on the characteristics of the underlying relationships, the rightsholders' relationship to the context of use, and the ways in which the digital transformation of acts of representation can and does impact their value.

Rooting the valuation of data in use has a number of important, secondary effects - importantly, it moves from a universal abstract directly into focusing on the value of the underlying decision, the impact of the data representation on the decision, and the contextualizing role of the interests representative. While the former two considerations are key for understanding the applied value of data in a given context, the last - the role and interests of the representative - are fundamental, and categorically undervalued, where not outright ignored, in mapping data supply chains and economies. Perhaps the most clarifying advantage of framing a data valuation through the context of representation is that it starts from the recognition that representations, especially those made on the behalf of others, require a legitimate basis. People are not entitled to represent you - especially in high-impact, rights-affecting contexts - simply because they purport to hold information about you. That reality is as true for the digital intermediaries and data brokers transforming supply chains in regulated fields as it is for the professionals they serve.

Data use as an act of representation

The exchange of data not only raises questions about the validity of the facts asserted, but also how and why the parties exchanging those facts are the appropriate actors to be doing so. If we ask these questions from the perspective of data production, suggesting that we try to regulate or limit who gets to know or share specific types of information is not only incredibly impractical, it runs afoul of a considerable number of globally recognized human rights. And starting from a context of use quickly brings us to the ludicrous position of considering whether a company that bought data from your home vacuum could provide that information against you in court or that your watch's activity monitor might get used to set your health insurance premiums. And yet– that is precisely the position in which we find ourselves (e.g. Aitken, 2017; Koebler, 2023) What is systemically lacking is a regularization across regulated spheres of the flow of information in high-impact settings. Historically, the mediator of these information flows has most commonly been situated in the role of regulated, fiduciary representatives.

It's perhaps most important to say here that the role of fiduciary representatives, even those whose services are primarily about the presentation of information - like lawyers – is not to present information. The role of a fiduciary representative is to represent another person's, or group of people's, interests in a defined context. And while there are lots of good and important arguments about the confluence and dissonance between an individual person's interests and the collective interest in objective truth or correctness - the role of a fiduciary is to be the expert advocate for whatever best serves their clients, whether it's treatment or honoring a 'do not resuscitate' order. (Though see Gold & Miller, 2015 for an exploratory synthesis of fiduciary duties to private and public interests.) The importance here, however, is that in order for data use to meet the standards that representatives need to meet - and representative's representations need to meet - the production supply chain not only needs to be explicit, it needs to be accessible. In other words, transparency is no substitute for governance (Gansky & McDonald, 2022).

In order to be a fiduciary representative, a professional must be able to understand their client's interests, triangulate the data and resources available to advance those interests, and be able to describe their representation to both the client and the decision-making context. In other words, in order for data to be suitable as a representation, by a representative, in a high-value context, it not only needs to be fit-for-purpose, but it has to come from a representative source with a legitimate basis to make that assertion.

There are, of course, a wide range of bases for making legitimate representations, even in high-impact settings, - but nearly all of them stem from one of three sources: (1) a representative relationship with the data subject's interests; (2) expertise that is directly relevant to the subject matter or interpreting information that is directly relevant to the subject matter; or (3) directly observed experience relevant to the decision. Each of these bases of legitimacy are implemented, governed, and held accountable in different ways, both procedurally and institutionally, based on the context of the decision made. For example, representatives that are acting in a person's interests typically have a direct relationship with the data subject, including direct contracting that sets expectations and outlines accountabilities. By contrast, a subject matter expert may not have any relationship with the data subject, but will be called upon - either by the people involved or by an adjudicating system - and has to be able to demonstrate objectivity, articulate the basis of their expertise, and stay within the limits of specific inputs. The practical impact of the contextual requirements for high-impact representations is that data users not only need data, they also need to be able to document and describe a diversity of legitimate bases for their acquisition and use of that data. And the responsibility for meeting those requirements, for both truth and legitimacy, falls directly and explicitly on the representative, typically imposed by the institutions that govern the context of use - many of whom have little, if any, relationship with the digital intermediaries who increasingly have a footprint in regulated contexts and services.

From speech to supply chains - the impact of digital transformation of representations

One of the primary differences between the existence of data and its use as a representation, and especially a digital representation, is that the latter acknowledges the context, intended impact, and the associated liabilities for that representation. Historically, this was actually a lot less complicated than it sounds. In many high-impact contexts, representations were required to be articulated in person (i.e. in court testimony) or through the use of a controlled proxy (such as a certified signature or wax seal). In these kinds of instances, representatives occupied a particular physical context which they, to some degree, perceived and communicated into - and so it makes sense to think they'd be able to be held accountable for things like the impact of the information they communicated, especially when done with intention. The affordances of digital technology, however, have significantly complicated this situation.

Data has not only created an explosion in the kinds of behaviors that can be observed, it has similarly complicated the number of actors and contexts implicated by the creation and use of data. The idea of a single person or organization making a representation may now just be the last link in a very long supply chain - one that includes data from a variety of sources, processed by a series of algorithms from different sources, and through tooling designed for an unknowable range of purposes.

The design of data and technology products, especially, obviate accountability through their legal affordances - each link in the supply chain described above, for example, is unique - and may operate in different jurisdictions, have made no representations or warranties about the fitness of their data or tool's creation for a specific purpose, and may not limit who can use it, or to what ends (S. McDonald, 2019; S. M. McDonald, 2019; Nissenbaum, 1996). Each of those fundamental protections are ways to ensure the integrity of representations and products - not just in the technical substance of the data they produce, but in ensuring that the actors involved had a legitimate basis to make the underlying representations. Starting from the perspective of data production, nearly every aspect of data creation, processing, sharing, and use complicates the basic mechanics of institutional accountability and contextual governance.

And yet, starting from the point of data use - especially in fiduciary relationships - contextual governance forces the representative to take singular responsibility for the appropriate use of data. Courts do not allow, for example, just any person to wander in and, with no relationship to the parties, the court, or the subject matter, make argumentation or submit evidence. That's not because we assume that people won't try, it's because the physical, procedural, and practical design of legal systems makes it difficult to do so. The integrity of the representatives and representations made in rights-affecting contexts is protected, but protected by the context of use - not by the supply chain of production. And the ways that data is produced, currently, are designed to avoid the common pathways of establishing agency, liability, and responsibility - often in ways that should disqualify them from use in high-value contexts. In other words, in the transition to data-driven decision-making, a number of important systems have gone from ensuring the contextual integrity of a representative pursuing a person's interests to trying to technocratically reverse engineer a universal or standardized quality assessment that justifies its use, regardless of the politics of its production (Keyes, 2021). And that has not only enabled a flood of new data, and would-be representatives, but it has fundamentally overwhelmed the capacity of the institutions responsible for protecting the integrity of decision-making contexts.

Said more simply, the data economy, in its current form, relies on its ability to brute-force the contextual, use-based, protections that predominantly define the requirements for governance as a necessary indicator of value. While there's no arguing that strategy's existence or prevalence, it is also easy to see the ways in which the erosion of integrity measures in high-impact, high-value contexts are a net negative - both for the number of people impacted now, and the ways in which the data economy is transforming system quality in the long-term. A counter-intuitive approach, then, is to focus on the means of defining and practically enforcing the governance relationships that go into high-integrity representation of another person's interests, as framework for assessing the value of data production, from the point of use. The most established, implemented, and available model for high-integity representations is that of the fiduciary.

Fiduciary Models of Representation and Governance in Valuation

The highest-importance decisions are the highest impact - and often, then, the most valuable (just look at how every major tech company has tried, and mostly failed, to enter medicine and medical informatics). Participation in those contexts is both contested and valuable - but not derived from the quality of a product, so much as the quality of a service delivered in a context - typically by a representative. So participation in high-value decisions, especially those influenced by data- and computation-intensive processes, is valuable and, increasingly, attempting valuation based on quality of process/product - as opposed to quality of impact/outcome. This misunderstands the purpose and value markers for a representation relationship (and, to a large extent, data's value and impact as a representation).

Using representation services as the model - and reverse engineering a valuation framework from a contextual decision made within that representation, then, necessarily has to have models for articulating, implementing, and enforcing, the quality of representative services. Fiduciary law and models are what most of the legal world uses to articulate, implement, and govern high-integrity, asymmetrical power relationships - essentially, relationships between people and those that represent their interests. Fiduciary models rely on three things: duties of care, duties of loyalty, and independent oversight.

Duties are different than standards - they require active, case-by-case consideration in ways that don't appeal to universalizable rules that abstract to the technical level/layer - but they do offer a functional model and set of system requirements that can be used to evaluate the integrity, quality, and thus, perhaps, value of a data supply chain as a representative relationship and data use as an act of representation.

<u>Representative Purpose Limitation + Defining and Reverse Engineering Value from Context</u>
Valuation constructs and derivations intentionally eschewed for scale by most digital tools, but data economies specifically - rely on a set of contextually derived factors that determine not only the value of the decision data is used to influence, but the character, quality, and effectiveness of the representative of each party's interests

Contextual definition (what's at stake)

Requires contextual interest mapping (who wants what in each situation)

Representational outcome mapping (what is possible and/or reasonable to achieve for the representative)

<u>Relationship Definitions and Bounding</u>

Scope of representation - time, decision-context, interests, assets, artifacts
Expertise, Performance, and Certification

<u>Direct and Independent Governance and Oversight</u>
Clarity on relevant authorities
Specific vs. holistic liabilities (process vs. outcome liabilities)
Reporting formats and identification/bounding of relevant information/factors

Conclusion

We can't regulate or valuate data as an object, but we can use digital transformation as a vector to understand both the characteristic requirements for data systems to deliver contextual value and a framework to understand the value of data's value in-context. Our ability to do so requires aligning our data valuation frameworks toward the quality and integrity indicators we use to create and situate the value of representations and their underlying relationships. Not only their interests, but the mapping, definition, and bounding of the constituent interests embedded in decision points is, at the granularity of data, a relatively novel economic, legal, and operational proposition. By contrast, the fiduciary model of defining and articulating representative relationship quality is at least instructive - but it points away from standardizable, universalizable valuations, toward granular, situated, and predominantly marginal understandings of data's value.

Ultimately the value of data is conditioned by the integrity, accessibility, and ongoing oversight of the supply chains which produce and mobilize data from their point of origin to their use as representations to influence decision-making. The particularities of what makes such supply chains fit for purpose is dependent on the context in which data as representations are articulated. Fiduciary models are by no means a universal or perfect solution (Khan & Pozen, 2019). However, the infrastructures through which fiduciaries are held accountable and their conduct is governed are useful for data worlds in two senses: as structural inspirations for architectures of continuous oversight, burdens of proof, and contestability for representations; and as existing professional regulations which can be more actively and vigorously applied to digitally-transformed acts of representation.

In this article we have proposed duties of care, duties of loyalty, and independent oversight as important terrains on which to advance professional standards with respect to digital transformation. For data and data systems to be helpful rather than harmful for data subjects and data users alike, rooting models of data value in integrity to purpose is an essential path forward.

Commented [1]: The argument is more simple than this - it's to say, in order for data to be fit for high-value purpose, it has to enable the people who make decisions in those contexts to meet their requirements they're held to.

We're making a valuation argument. The argument is that these things are part of value.

Sean Martin McDonald is a Director at Digital Public and Senior Fellow at the Centre for International Governance Innovation.

Ben Gansky is a PhD candidate at Arizona State University's School for the Future of Innovation and Society.

Works Cited

Aitken, R. (2017). 'All data is credit data': Constituting the unbanked. *Competition & Change*, *21*(4), 274–300. https://doi.org/10.1177/1024529417712830

Almklov, P., Østerlie, T., & Haavik, T. (2014). Situated with Infrastructures: Interactivity and Entanglement in Sensor Data Interpretation. *Journal of the Association for Information Systems*, *15*(5), 263–286. https://doi.org/10.17705/1jais.00361

Aronova, E., Oertzen, C. von, & Sepkoski, D. (2017). Introduction: Historicizing Big Data. *Osiris*, *32*(1), 1–17. https://doi.org/10.1086/693399

Baker, K. S., & Millerand, F. (2007). Articulation Work Supporting Information Infrastructure Design: Coordination, Categorization, and Assessment in Practice. *2007 40th Annual Hawaii International Conference on System Sciences (HICSS'07)*, 242a–242a. https://doi.org/10.1109/HICSS.2007.88

Balkin, J. M. (2020). *The Fiduciary Model of Privacy* (SSRN Scholarly Paper 3700087). https://papers.ssrn.com/abstract=3700087

Bates, J. (2014). Climate Risk, Big Data and the Weather Market. *SPERI*, *13*.

Bates, J., Lin, Y.-W., & Goodale, P. (2016). Data journeys: Capturing the socio-material constitution of data objects and flows. *Big Data & Society*, *3*(2), 205395171665450. https://doi.org/10.1177/2053951716654502

Beauvisage, T., & Mellet, K. (2020). Datassets: Assetizing and Marketizing Personal Data. In K. Birch & F. Muniesa (Eds.), *Assetization: Turning Things into Assets in Technoscientific Capitalism*. The MIT Press. https://doi.org/10.7551/mitpress/12075.001.0001

Benjamin, R. (2019). *Race After Technology: Abolitionist Tools for the New Jim Code*. Polity

Press.

Birch, K., Cochrane, D., & Ward, C. (2021). Data as asset? The measurement, governance, and valuation of digital personal data by Big Tech. *Big Data & Society*, *8*(1), 20539517211017308. https://doi.org/10.1177/20539517211017308

Bogen, J., & Woodward, J. (1988). Saving the Phenomena. *The Philosophical Review*, *97*(3), 303. https://doi.org/10.2307/2185445

Bokulich, A. (2020). Towards a Taxonomy of the Model-Ladenness of Data. *Philosophy of Science*, *87*(5), 793–806. https://doi.org/10.1086/710516

Bokulich, A., & Parker, W. (2021). Data models, representation and adequacy-for-purpose. *European Journal for Philosophy of Science*, *11*(1), 31. https://doi.org/10.1007/s13194-020-00345-2

Borgman, C. L., Scharnhorst, A., & Golshan, M. S. (2019). Digital data archives as knowledge infrastructures: Mediating data sharing and reuse. *Journal of the Association for Information Science and Technology*, *70*(8), 888–904. https://doi.org/10.1002/asi.24172

Bowker, G. C. (2005). *Memory practices in the sciences*. MIT Press.

Bowker, G., & Millerand, F. (2008). Metadata Standards: Trajectories and Enactment in the Life of a Standard. In M. Lampland & S. L. Star (Eds.), *Standards and Their Stories: How Quantifying, Classifying, and Formalizing Practices Shape Everyday Life* (1st edition). Cornell University Press.

Bowker, G., & Star, S. L. (1999). *Sorting things out: Classification and its consequences*. MIT Press.

Braun, L. (2014). *Breathing Race into the Machine: The Surprising Career of the Spirometer from Plantation to Genetics* (1st edition). Univ Of Minnesota Press.

Chun, W. H. K., & Barnett, A. (2021). *Discriminating data: Correlation, neighborhoods, and the new politics of recognition*. The MIT Press.

Collington, R. (2019). *Digital Public Assets: Rethinking value and ownership of public sector* *data in the platform age*. Common Wealth.

Regulation of the European Parliament and of the Council on European Production and Preservation Orders for electronic evidence in criminal proceedings and for the execution of custodial sentences following criminal proceedings. Retrieved January 25, 2023, from https://data.consilium.europa.eu/doc/document/ST-5448-2023-INIT/en/pdf

Crain, M. (2018). The limits of transparency: Data brokers and commodification. *New Media & Society*, *20*(1), 88–104. https://doi.org/10.1177/1461444816657096

D'Ignazio, C., & Klein, L. F. (2020). *Data Feminism*. The MIT Press.

Doctorow, C. (2008, January 15). Personal data is as hot as nuclear waste. *The Guardian*. https://www.theguardian.com/technology/2008/jan/15/data.security

Edwards, P. N. (2010). *A vast machine: Computer models, climate data, and the politics of global warming*. MIT Press.

Edwards, P. N., Mayernik, M. S., Batcheller, A. L., Bowker, G. C., & Borgman, C. L. (2011). Science friction: Data, metadata, and collaboration. *Social Studies of Science*, *41*(5), 667–690. https://doi.org/10.1177/0306312711413314

Ferryman, K. (2017). *Reframing Data as a Gift* (SSRN Scholarly Paper 3000631). https://doi.org/10.2139/ssrn.3000631

Gansky, B., & McDonald, S. (2022). CounterFAccTual: How FAccT Undermines Its Organizing Principles. *2022 ACM Conference on Fairness, Accountability, and Transparency*, 1982–1992. https://doi.org/10.1145/3531146.3533241

Gitelman, L. (Ed.). (2013). *Raw Data Is an Oxymoron*. MIT Press. https://doi.org/10.7551/mitpress/9302.001.0001

Gold, A. S., & Miller, P. B. (2015). Fiduciary Governance. *William and Mary Law Review*, *57*.

Gressin, S. (2023, September 11). *FTC lawsuit insists on FCRA compliance and transparency from background report providers*. Federal Trade Commission. https://www.ftc.gov/business-guidance/blog/2023/09/ftc-lawsuit-insists-fcra-compliance-transparency-background-report-providers

Hussain, M. I., & Bowker, G. C. (2021). Phenotyping as disciplinary practice: Data infrastructure and the interprofessional conflict over drug use in California. *Big Data & Society*, *8*(2), 20539517211031258. https://doi.org/10.1177/20539517211031258

Hwang, T. (2020). *Subprime Attention Crisis*. FSG.

Igo, S. E. (2018). Me and My Data. *Historical Studies in the Natural Sciences*, *48*(5), 616–626. https://doi.org/10.1525/hsns.2018.48.5.616

Katyal, S. K. (2019). THE PARADOX OF SOURCE CODE SECRECY. *CORNELL LAW REVIEW*, *104*.

Keyes, O. (2021). Standard Evasions. *Real Life*. https://reallifemag.com/standard-evasions/

Khan, L. M., & Pozen, D. E. (2019). A SKEPTICAL VIEW OF INFORMATION FIDUCIARIES. *Harvard Law Review*, *133*.

Knorr-Cetina, K. (1999). *Epistemic cultures: How the sciences make knowledge*. Harvard University Press.

Koch, B., Denton, E., Hanna, A., & Foster, J. G. (n.d.). *Reduced, Reused and Recycled: The Life of a Dataset in Machine Learning Research*.

Koebler, J. (2023, September 28). *Food Delivery Robots Are Feeding Camera Footage to the LAPD, Internal Emails Show*. 404 Media. https://www.404media.co/serve-food-delivery-robots-are-feeding-camera-footage-to-the-lapd-internal-emails-show/

Lauer, J. (2017). *Creditworthy: A History of Consumer Surveillance and Financial Identity in America* (p. 368 Pages). Columbia University Press.

Lemov, R. (2015). *Database of Dreams: The Lost Quest to Catalog Humanity*. Yale University Press.

Leonelli, S. (2015). What Counts as Scientific Data? A Relational Framework. *Philosophy of Science*, *82*(5), 810–821. https://doi.org/10.1086/684083

Leonelli, S. (2019). Data Governance is Key to Interpretation: Reconceptualizing Data in Data

Science. *Harvard Data Science Review*. https://doi.org/10.1162/99608f92.17405bb6

Leonelli, S., & Tempini, N. (Eds.). (2020). *Data Journeys in the Sciences*. Springer International Publishing. https://doi.org/10.1007/978-3-030-37177-7

Lightbourne, J. (2017). Damned Lies & Criminal Sentencing Using Evidence-Based Tools. *Duke Law & Technology Review*, *15*(1).

McDonald, S. (2019). *From Space to Supply Chains: A Plan for Humanitarian Data Governance* (SSRN Scholarly Paper 3436179). https://doi.org/10.2139/ssrn.3436179

McDonald, S. M. (2019). *The Fiduciary Supply Chain*. Centre for International Governance Innovation. https://www.cigionline.org/articles/fiduciary-supply-chain/

McGuigan, L. (2019). Automating the audience commodity: The unacknowledged ancestry of programmatic advertising. *New Media & Society*, *21*(11–12), 2366–2385. https://doi.org/10.1177/1461444819846449

McKelvey, F. (2022). When the New Magic was New: The Claritas Corporation and the Clustering of America. *IEEE Annals of the History of Computing*, *44*(4), 44–56. https://doi.org/10.1109/MAHC.2022.3214223

McNealy, J. E. (2021). *An Ecological Approach to Data Governance* (SSRN Scholarly Paper 4164112). https://doi.org/10.2139/ssrn.4164112

Morgan, M. S. (2020). The Datum in Context: Measuring Frameworks, Data Series and the Journeys of Individual Datums. In S. Leonelli & N. Tempini (Eds.), *Data Journeys in the Sciences*. Springer International Publishing. https://doi.org/10.1007/978-3-030-37177-7

Nissenbaum, H. (1996). Accountability in a computerized society. *Science and Engineering Ethics*, *2*(1), 25–42. https://doi.org/10.1007/BF02639315

Nissenbaum, H. (2009). *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Stanford University Press.

Obar, J. A. (2020). Sunlight alone is not a disinfectant: Consent and the futility of opening Big Data black boxes (without assistance). *Big Data & Society*, *7*(1), 205395172093561.

https://doi.org/10.1177/2053951720935615

O'Reilly, T. (2021). *Data Is the New Sand*. The Information. https://www.theinformation.com/articles/data-is-the-new-sand

Ottinger, G. (2010). Epistemic Fencelines: Air Monitoring Instruments and Expert-Resident Boundaries. *Spontaneous Generations: A Journal for the History and Philosophy of Science*, *3*(1), 55–67. https://doi.org/10.4245/sponge.v3i1.6115

Parsons, A., & Viljoen, S. (2023). *Valuing Social Data* (SSRN Scholarly Paper 4513235). https://doi.org/10.2139/ssrn.4513235

Pine, K. H., & Liboiron, M. (2015). The Politics of Measurement and Action. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 3147–3156. https://doi.org/10.1145/2702123.2702298

Rappert, B., Wilson-Kovacs, D., Wheat, H., & Leonelli, S. (2022). Evincing Offence: How Digital Forensics Turns Big Data into Evidence for Policing Sexual Abuse. *Engaging Science, Technology, and Society*, *8*(3), Article 3. https://doi.org/10.17351/ests2022.1049

Richards, N. M., & Hartzog, W. (2015). Taking Trust Seriously in Privacy Law. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2655719

Richards, N. M., & Hartzog, W. (2020). A Duty of Loyalty for Privacy Law. *Washington University Law Review*, *99*. https://doi.org/10.2139/ssrn.3642217

Sadowski, J. (2019). When data is capital: Datafication, accumulation, and extraction. *Big Data & Society*, *6*(1), 205395171882054. https://doi.org/10.1177/2053951718820549

Tempini, N. (2020). The Reuse of Digital Computer Data: Transformation, Recombination and Generation of Data Mixes in Big Data Science. In S. Leonelli & N. Tempini (Eds.), *Data Journeys in the Sciences*. Springer International Publishing. https://doi.org/10.1007/978-3-030-37177-7

Thylstrup, N. B., Hansen, K. B., Flyverbom, M., & Amoore, L. (2022). Politics of data reuse in machine learning systems: Theorizing reuse entanglements. *Big Data & Society*, *9*(2), 205395172211397. https://doi.org/10.1177/20539517221139785

Turow, J. (2012). *The Daily You: How the New Advertising Industry Is Defining Your Identity and Your Worth*. Yale University Press.

Turow, J. (2017). *The Aisles Have Eyes: How Retailers Track Your Shopping, Strip Your Privacy, and Define Your Power*. Yale University Press.

Vertesi, J., & Dourish, P. (2011). The value of data: Considering the context of production in data economies. *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work - CSCW '11*, 533. https://doi.org/10.1145/1958824.1958906

Viljoen, S. (2021). A Relational Theory of Data Governance. *The Yale Law Journal*.

Wexler, R. (2017). Life, Liberty, and Trade Secrets: Intellectual Property in the Criminal Justice System. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2920883

Zuboff, S. (2015). Big other: Surveillance Capitalism and the Prospects of an Information Civilization. *Journal of Information Technology*, *30*(1), 75–89. https://doi.org/10.1057/jit.2015.5