IARIW – CIGI 2023

IARIW – CIGI 2023

Thursday, November 2 – Friday, November 3

Be Careful What You Wish Measure (For)
An Onto-Epistemological Fable of ('Personal') Data Between Web 2.0 and Web 3.0

Georgia Meyer
(London School of Economics and Political Science)

Paper prepared for the Conference on The Valuation of Data November 2 – November 3, 2023
Session 7: Personal and Platform Data
Time: Friday, November 3, 2023 [11:00AM-12:00PM EST]

IARIW-CIGI Conference
'The Valuation of Data'

Be Careful What You ~~Wish~~ Measure (For)
An Onto-Epistemological Fable of ('Personal') Data Between Web 2.0 and Web 3.0

Georgia Meyer

Abstract

Emerging ways to store and use digital 'personal' data promise to decouple data from identity and applications - ending the centralising patterns of data generation and accumulation that have characterised 'Web 2.0' and entrenched particular ways to define and generate value. In a new vision for data relations, 'Web 3.0', individuals would have more 'control' over 'their' data but it is not yet clear what this will mean in practice. Can data belong to individuals when it has relational impacts (both positive and negative)? Can greater control be realised in more granular choices about granting and revoking access to data? Perhaps a deeper interrogation of how the ways data is created, measured and used, comes to structure realities is required.

I. Introduction

It is well understood that the past two decades of the World Wide Web has been characterised by the collection of both vast swathes of people's online behaviours as digital (personal) data and a proliferation of user-generated content (UGC). This period, known as 'Web 2.0', characterised by interactivity between individual 'users' mediated by platforms and their applications, that collect and store this data in centralised, siloed fashion, has run alongside an era of 'Big Data' (Birch et al. 2021; boyd & Crawford 2012; Kitchin 2014; Tene & Polonetsky 2013). This phenomena, though not entirely a product of the Web 2.0 (since it comprises an expanded set of digital data collection devices and activities beyond web-based platforms and applications), nonetheless shares some key traits in that it has been a largely centralised and centralising force of a digital 'data imperative' (Fourcade & Healy 2016) to collect (/create) as much data as possible. The commercial justification underpinning this drive was that such data would generate value 'downstream' (Tene & Polonetsky 2013).

Research has demonstrated how these amassed varieties of digital personal data are transformed into assets legible to various business, market, and financial logics (Birch et al. 2021) as well as underpin 'prediction products' (Zuboff 2019). That these processes of 'techcraft' have delivered astronomical returns to technology companies is also well noted (Birch et al. 2021). Though what has been more elusive is measuring the utility gains for users of free services and accounting for their relationship with the wider economy (Brynjolfsson et al. 2019). From the perspective of public policy, personal data takes on a different guise and comprises essential information to gain an understanding of, and develop policy that intervenes, on all manner of questions related to health, transport, sustainability, resource allocation, education and beyond. Efforts are underway to develop methods to 'measure' such (potential) value in order to justify the costs in making such data available (Coyle & Diepeveen 2021).

Regulatory responses to this environment of commercial data hoarding have been data protection legislation, underpinned by principles of information privacy, which have sought to regulate data flows based on defining and demarcating 'personal data'. Opportunities for individual level consent have been introduced, but these have been widely criticised as too narrow and ultimately meaningless (Bergemann 2018; Carolan 2016; Mantelero 2014). Concurrently, interdisciplinary theoretical debates in academia have been (re)articulating normative aspects of information privacy (Knijnenburg et al. 2022; Smith et al 2011; Solove 2007), whilst empirical contributions have developed and tested theory-informed models to understand people's online data sharing behaviours (Dinev et al. 2015; Li 2011; Malhotra et al. 2004; Smith et al. 1996; Smith 2004; Smith et al. 2011), particularly when behaviours run contrary to stated concerns about privacy and data sharing as in the so-called *Privacy Paradox* (Belanger et al. 2002; Norberg et al 2007), or when exchanging privacy for short term utility gains as in the *Privacy Calculus* (Acquisti & Grossklags 2005; Acquisti et al. 2015; Acquisti & Wagman 2016).

As regulatory responses to this widely critiqued data ecosystem have been moving from regulating data flows to regulating harmful inference from data (Crawford 2013; Tene & Polonetsky 2013; Xu & Dinev 2022), technologies have also been in development that could provide some remedies. One particular approach, known as 'Web 3.0' (not to be confused with blockchain based 'Web3' technologies), is a movement to give people more 'control' over 'their' data, using a set of semantic web protocols and standards that provide the technological architecture for a web where identity, data, and applications are decoupled. Rather than digital data being stored by applications in siloed fashion, data is stored around entities - like individuals - in a Personal Data Store (PDS) comprised of various pods (e.g. for an individual's government, health, media data). The basic idea is that the individuals to whom these pods are assigned are provided with a range of choices about what level of access to grant to 'their' data, to what entity, and for how long (Inrupt 2023a, 2023b).

In contrast to the choices presented to people in, for example, cookie consent forms, these opportunities for decision-making about data sharing in PDS would not presented in ad hoc fashion and at inopportune and inconvenient moments. The Web 3.0 approach could provide opportunities for deliberative decision-making about granting and revoking access to data in relation to enterprise and government data use - building on ideas of Dynamic Consent (Kaye et al. 2015). Alongside the Web 3.0 approach various civil society actors, policymakers and academics are developing frameworks for bottom up approaches to data use and (re)use that are being designed to give those from whence data came more meaningful input (Aapti Institute 2021; Aapti Institute & Open Data Institute 2021; Manohar 2020a, 2020b; Mozilla Foundation 2020a, 2020b; Mozilla Foundation 2022; Nanda 2021; Ramesh & Kapoor 2020).

Those concerned about a wildly out of balance political economy of data (Zuboff 2015) may find this scenario appealing. But there are significant questions about what such an alternative-universe of data relations rendered through technology could actually deliver should it begin to materialise. These go to the very heart of ontological questions about the nature of 'personal' data - also bound up with questions of 'its' value. Scholarship has pointed out that conceiving of data as something that can be 'owned' by an individual and traded according to an economic and legal lexicon of property rights and market valuations aren't appropriate for data 'given their relational characteristics' (Coyle 2022) and other market defying properties. Beyond this, since the value (and protection) of personal data is bound up with both individual concerns and wider societal, environmental and economic costs and benefits, is a framework of protection (informed by normative ideas about privacy), or governance, based on 'data as an individual medium' (Viljoen 2021) sufficient?

Finally, if we consider data from the vantage point of how it is bound with the co-constitution of reality (not 'just' 'value'). This underlines the question of what opportunities people are being given to participate in the constructions of 'their' (/our) data futures. The creation, collection and deployment of ('personal') data is linked to decisions over what to make salient - what to make legible in the first place. Beyond bringing legibility, digitally collected and processed data has become fused with sophisticated computational modelling techniques designed for the prediction and simulation of the future decisions and behaviours of people in complex societies and economies. Thus not only is it the engine of downstream insight generation and the foundation for a 'new knowledge economy' (Pentland et al. 2021: 35), it is also a window (and bridge) to the future(s) through processes of salience making as well as sketching the contours of ontological imagination. With this in mind then perhaps neither consent (as in Web 2.0), nor granting access to uses (as in Web 3.0), may provide data 'producers' an appropriate level of involvement in conceiving its use.

This short essay attempts to bring some of these questions, a subject of much conversation, into analytical focus using the concepts of identifiability, control and value across Web 2.0 and Web 3.0 data ecosystems. Particularly to lay some groundwork for considering whether greater dexterity in access control mechanisms could be a supplementary means to meaningfully restore balance in the emerging economy of data relations.

II. Data Between Web 2.0 and Web 3.0

Academic discussions concerned with privacy and data governance in the Information Systems (IS) field, as well as discussions about the Economics of Privacy from a broader range of scholars, have run in parallel with the Web 2.0 era. It is possible that technologies that seek to redraw the architectures of data storage could represent a distinct era and that this emerging empirical context could precipitate new conceptualisations of 'personal' data and value as well as new methodological opportunities. The next section briefly covers the nature of the data ecosystems in Web 2.0 and Web 3.0 and touches on implications of knowledge-making.

II.I. Characteristics of Web 2.0

II.I.i. What data?

Web 2.0 in the form of 'blogs, social media, online social networks' has 'rendered individuals no longer mere consumers of information but public producers of often highly personal data' (Acquisti et al 2016: 444). This 'so called Web 2.0' (Acquisti et al 2016) thus refers to this era of the web typified by the rise of user-generated content (UGC), facilitated by the rapid global rise of social media companies and mobile devices with concurrent developments in business models and economic models via e-commerce (Li 2011) and digital platforms (McIntyre et al. 2021; Rochet & Tirole 2003) which have spawned entire new digital ecosystems (Ghazawneh & Henfridsson 2015).

Cumulatively these developments constitute the era of Big Data (Kitchin 2014), often largely what is known as digital trace data (Aaltonen & Stelmaszak 2023; Berente et al. 2019; Power 2022). Consequentially, shifts in research approach and agendas have emerged (Abbasi et al. 2016) and the training data fuel of Web 2.0 helped precipitate the paradigm-shifting developments in generative AI, most notably the suite of GPT models of the last few years.

II.I.ii. What knowledge?

Concurrently there has been a scholarly debate on the nature of knowledge production that accompanied the arrival of Big Data. Some, especially in business and technology circles, have described 'a new era of empiricism', whereby the sheer volume of data offered ways to mine data to reveal patterns that would not have been investigated if informed by theory (Kitchin 2014). If not handled with caution, there is an implicit assumption within this perspective, that the data at hand is an accurate and complete enough picture of 'reality' to claim that any such conclusions drawn about patterns derived from data mining are valid counterparts to our understanding of reality - rather than being mere models of (digital) Model Lands (Thompson 2022) of (human) digital trace data. This is significant since it has implications for how the boundaries of ontological imagination are drawn, and by whom, and what it is that people might/ought to seek control over when and if they are provided with such opportunities.

II.II. Characteristics of Web 3.0

II.II.i. What data?

The suite of technologies that together constitute Web 3.0 are decentralised databases and Personal Data Stores (PDS) supported by a set of protocols and languages (the "semantic web") such as Resource Description Framework (RDF) and Web Ontology Language (OWL) (Inrupt

2023a, 2023b, 2023c, 2023d). Though some of these technologies have been in development for over twenty years their adoption in practice has only been made possible more recently with trial launches of PDS. Solid was launched in 2016 at MIT though computer scientists have been working on semantic web protocols since before Sir Tim Berners-Lee published an outline of his vision for a web of linked data (Berners-Lee & Fischetti 1999). For example, the first note on a proposed way to embed metadata in HTML was published by the World Wide Web Consortium (W3) in 1998 (World Wide Web Consortium 1998). The first presentation of Web Ontology Language (OWL) was published five years later (World Wide Web Consortium 2003).

All kinds of data can be stored in pods - covering both knowingly created data from users (e.g. photos, text, search queries) and less consciously created data (e.g. behavioural patterns like clicks, scrolls and dwell times, location and the recording of preferences). Certain kinds of data will be subject to specific regulations (e.g. health and financial data) and so can be stored in pods specific to those types of data. These pods are manageable by users in ways that comply with regulatory obligations and grant different levels of access controls to their owners. The Solid interoperable data standard is underpinned by the context-rich format in which the data is stored enabling multiple applications to work with the data should they be granted access (Inrupt 2023a, 2023b).

II.II.ii. What knowledge?

The key difference between a Web 2.0 and Web 3.0 data ecosystem is the level of control that producers of data may be granted over the uses of their data. This is by virtue of the standards, protocols and storage outlined above. However much of the control over data uses will still be managed through data protection regulations of which users granting access to is simply one component. There is however an adjacent aspect of data use and knowledge creation worth considering here given the wave of research and exploration of the potential of Big Data for unlocking value and the inherent tensions between innovation and protection that lie at the heart of this trend.

Kitchin (2014) has discussed that an evolution of naive big data empiricism, that might work extremely well for commercial objectives, has been replaced by more conscientious data-driven epistemologies. Such epistemologies retain the value of domain expertise and theory enhanced exploration whilst similarly being open to the exploratory opportunities offered by new and richer forms of data. If understandings of data's relational characteristics and its multidimensional values are to be reflected in decentralised data ecosystems it is interesting to consider what new forms of knowledge might result. A greater emphasis on contingencies, exploration, possibilities and emergence could be required to make sense of the shifting landscape (Arthur 2020). If embarked upon with greater input from data 'producers' perhaps some of the inherent tension between innovation and protection can be softened.

III. Identifiability, Control & Value

To operationalise how to consider any key differences between Web 2.0 and Web 3.0 practices that have relevance for understanding emerging/changing data relations I suggest the use of three concepts: identifiability, control and value. Setting them side by side it is possible to observe not just that there might be changes in their materialisation across Web 2.0 and Web 3.0 but that there might be changing contingencies between the various components of identifiability, control and value (as/if those components are realised differently in new empirical contexts).

The reasons I have selected these three concepts are due to the centrality of identifiability and control (Xu 2007) in literatures to date on information privacy research and the expanded interest in the value of information as an adjacent academic conversation. **Table 1** sets out the questions I asked when reviewing how these concepts have been discussed, if at all, in literatures about data sharing and value nascent to Web 2.0.

	Identification	Control	Value
Operationalisation of key concepts for theoretical literature review	• How is the question of whether the information being disclosed is 'personally identifiable' addressed in the studies? • How do the studies interrogate or operationalise ideas about 'inferred knowledge' or 'aggregate data'? • How do the studies attempt any kind of classification of different types of 'personal information' being disclosed?	• How is the concept of control operationalised in the studies (if at all)? • If operationalised, is control discussed in relation to: (i) whether to disclose; (ii) level of disclosure; (iii) what to disclose; or, a combination? • How is control operationalised in terms of control over the purposes to which data is put (if at all)?	• How is the concept of value operationalised, if at all, and through what conceptual proxies? • How is value discussed in relation to individuals and / or societies and / or commercial entities? • What is the measuring agent of value in the studies i.e is it an individual making the assessment of value and / or is value (or risks) defined by other entities other than the individual making the data sharing decision?

Table 1. Structured approach for literature review

Table 2 sets out key literatures and theoretical orientation of research into information privacy and value during the Web 2.0 and big data periods. These are discussed in greater length in the following sections by analysing where and how these research contributions, born of Web 2.0, touch on questions related to identifiability, control and value and their contingencies, if any. Each subsection notes some emerging questions for Web 3.0 based on the prior discussion. **Table 3** goes on to summarise some of the main differences that could emerge between Web 3.0 and Web 2.0 with regard to these concepts specifically.

	Empirical context	Research on privacy	Research on value
Web 2.0	• Empirical contexts that have supported and / or yielded subsidiary and adjacent theoretical devices to information privacy and the value of information have been ones in which data is sent to centralised databases - where it is stored by the collecting (commercial) organisation. These have been overwhelmingly e-commerce, social media and web search contexts. • Opportunities for individual level participation have been restricted to perfunctory consent opportunities or none at all.	• Privacy as a Right (Caudill & Murphy 2000; Margulis 1977a; Margulis 2003a). • Taxonomy of Privacy (Solove 2006). • Privacy as a function of control (Whitley 2009; Xu 2007). • Privacy concerns measured in the Internet Users' Information Privacy Concerns (IUIPC) model (Malhotra et al. 2004). • Antecedents Privacy Concerns Outcomes (APCO) macro model (Smith et al. 2011; Dinev et al. 2015): information privacy operationalised in various ways and combined into central construct of 'privacy concerns'. • Privacy concerns as a dependent variable (DV) in relation to factors such as personality differences, privacy experiences, demographic differences, culture/climate, privacy awareness (Dinev et al. 2015). • Privacy concerns as an independent variable (IV) observing effects on levels of trust, willingness to engage in commercial activities and other behavioural reactions (Dinev et al. 2015). • Privacy as a Commodity and the Economics of Privacy (Acquisti et al. 2016). • Privacy Calculus' and 'Privacy Paradox' (Acquisti et al. 2016; Belanger et al. 2002; Norberg et al 2007). • Recognition of individual level harms/ benefits and aggregated information harms/ benefits (Acquisti et al. 2016). • Identifying and theorising relational harms (Birhane & Cummins 2019; Crawford 2013; Viljoen 2021).	• Assetisation of personal data via 'techcraft' (Birch et al. 2021). • Commodification of attention via 'prediction products' (Zuboff 2019). • The 'data imperative' to collect as much data as possible (Fourcade & Healy 2016). • 'Big data empiricism' (Kitchin 2014). • The question of measuring the value of 'free goods' in the digital economy and the proposition of GDP-B (Brynjolfsson et al. 2019; Collis 2020). • Web scraping that supports the training of large-language models (LLMs) and other generative AI models (Bender et al. 2021).

Table 2. Theoretical concepts and empirical contexts of Web 2.0

III.I. Identifiability

III.I.i. Identifiability in Web 2.0

In the *Privacy as a Right* literatures personal information is treated as a phenomena inextricably linked with an individual's personhood and, since information privacy scholarship grew out of privacy scholarship more broadly, any threats to that personhood via unauthorised access to that information is deemed a risk. Solove's Taxonomy of Privacy (2006) is one of the most widely cited articulations of the various risks that unwarranted use of individual's information can pose to an individual. Discussion of aggregation is undertaken by Solove (2006: 505-512) and is framed as in relation to different pieces of information held about an individual on different databases which, when combined, would point to an identifiable individual. The key set of contingencies in this body of work is that personal information concerns identifiable individuals who ought to exercise control to avoid risks (whether provided by legal infrastructure or their own behaviours) that result from unauthorised or unplanned access to that information.

As literatures have argued, the increased sophistication and adoption of Privacy Enhancing Technologies (PETs) alongside advances in AI/ML are calling into question the extent to which 'My privacy is still about my private information' (Xu & Dinev 2022) and extended it to the question of inferred knowledge from large centralised data collections. Yet Solove's taxonomy does not refer to aggregation as a function of profiling and predictive harms where because someone has a particular characteristic they are algorithmically implicated in a decision, or insight, or advertising target by virtue of that characteristic (whether demographic, psychographic or behavioural with a different spectrum of harms attached to each as ample research has sought to draw attention to). This failure to recognise algorithmic harms (Birhane & Cummins 2019; Crawford 2013), born from data's relational nature materialised in aggregate processing, is now a catalyst for conversations to reconsider legal frameworks based upon the premise of 'Data as an Individual Medium (DIM)' (Viljoen 2021).

III.I.ii. Questions for identifiability in Web 3.0

The conversation about algorithmic harms reflects the loosening of our conceptions of data as something that is wedded to identifiable individuals to something that has relational impacts (both positive and negative) in the world. With this in mind we might also want to consider what a framework for data as a relational resource (not just a source of real, potential relational harm) could be, from the perspective of an *individual's* decision-making capacities (in collaboration with data scientists, economists and policymakers). Research that enhances routes to 'meaningfully informed consent' is taking place (Gomez Ortega et al. 2023), and progress in bottom up data institutions, facilitated by civil society groups is also underway (Aapti Institute 2021; Aapti Institute & Open Data Institute 2021; Manohar 2020a, 2020b; Mozilla Foundation 2020a, 2020b; Mozilla Foundation 2022; Nanda 2021; Ramesh & Kapoor 2020).

Other conversations are shifting understandings of individuals as merely passive 'consenters' to the uses of their data. Some are reframing data subjects as data 'sovereigns' who will be able to collectively bargain for their interests through data unions that monetise data on their behalf (Hill 2021). Other work is underway to craft a framework for 'digital self-determination' (Verhulst 2022). In this work, three asymmetries have been identified which are limiting 'how wider access to data can lead to positive and often dramatic social transformation': data asymmetries, information asymmetries, agency asymmetries (Verhulst 2022: 1). The emphasis here is not on reforming consent, or finding paths to individual or collective ownership but rather, as in the case of other civil society and advocacy group mobilisation, to design institutions, practices, licenses and codes of conduct that can steward data from individuals for the public interest (Verhulst 2022: 17-18). It may be that there is a role for emerging technologies to supplement and catalyse these efforts.

These various activities capture questions of whether value is about monetisation or about realising social value. They also demonstrate the complexities of theorising something that is at once bound with individuals and realised in unforeseeable ways far beyond them. Recognising that even though data comes from individuals, if people are given more opportunities to conceive of its use beyond themselves (not all of which are harmful), potentially fruitful ways to conceive of our complex relational responsibilities towards one another could emerge. As such it will be interesting to explore how people grant access to 'their data' in new empirical territory, like Web 3.0, and whether Web 2.0 nascent research designs for exploring data sharing, like those underpinning the Economics of Privacy, are still appropriate.

III.II. Control

III.II.i. Control in Web 2.0

Control is another key concept in articulations of information privacy to date, and conversely a feature that a loss of 'it' (a loss of control) has been lamented as a feature Web 2.0's omnipresent patterns of data accumulation and centralisation. Granting and maintaining consumer control of such data has long been perceived as an important feature of information privacy (Margulis 1977a, 1977b, 2003a, 2003b; Whitley 2009), even as it has also long been recognised as difficult to define (Whitley 2009). It has become even more complex to define as AI/ML techniques have become more sophisticated and knowledge generation more spatially and temporally disparate. This raises the question of what exactly it is control over that should be sought.

Opportunities for some kind of control (whatever that really means), at the level of the individual, has tended to be through the mechanism of consent, the limitations of which numerous contributions have drawn attention to. Key aspects of consent are that it must be *informed* consent and *freely given* (Bergemann 2018). However numerous critiques have raised the question of whether such ambitions can be met in practice in digital landscapes. On the first point, can consent really be informed when a) people often do not read privacy policies; b) if they do they are difficult to understand - 'the transparency paradox' (Barocas & Nissenbaum *in* Bergemann 2018: 116); c) decision-making about online consent is often 'skewed', likely haphazard, and in situations where users are often influenced 'by what their perceived short-term gains are' (Bergemann 2018: 116). On the question of consent that is freely given, critiques have pointed out the lack of genuine choice in monopolistic platform markets, users' increasing reliance on such platforms and that there is no real freedom in a 'take it or leave it deal' (Bergemann 2018: 115).

Additional controls over the uses of personal data are present at the structural level of legal protections which limit purposes, secondary uses of data and the volumes of data that can be collected. But these are hard to enforce and they still tend to render individuals somewhat passive in the economy of data relations even as the enactment of such principles into law have represented genuine milestones in attempts to reconfigure balances of power.

III.II.ii. Questions for control in Web 3.0

Web 3.0 ecosystems may be being built with an explicit ambition to return control to individuals, via more intricate access controls and different database architectures, but it is as yet unclear how such ecosystems can manage ongoing insight generation from knowledge ("information") even when access to that information was granted for a limited period (as may be the case in a PDS arrangement). Moreover, it is not clear how ideas about control might be augmented by the changing nature of individual identifiability in data sharing decisions, or changing ideas about value and data. It is worth exploring, therefore, whether the nature of control in new data ecosystems, could be shifting from solely thinking about control over information that is believed to be of intrinsic risk/value to individuals, and to control over the insight generating processes that give data value beyond said individual and in relation to societies more broadly.

III.III. Value

III.III.i.Value in Web 2.0

Research has demonstrated how various kinds of digital personal data are transformed into 'assets' which are legible to the business and financial logics that underpin value generation/extraction symbiotic with the lexicons of value in those contexts. Various practices of 'techcraft' (Birch et al) have been identified which comprise the classification of persons into 'users' and personal data into 'user engagement' (Birch et al 2021: 13). These 'techno-economic' objects, made measurable through metrics like Daily Active Users (DAUs) and Monthly Active Users (MAUs), are taken as a proxy for future monetisation. By shaping estimations of future customer base size, and future customer behaviours, these metrics can inform revenue forecasts from advertising (and/or subscriptions) which in turn mobilise venture capital and financial markets. Measuring 'both users and use reflects the techcraft that both creates *and* controls user data.' (Birch et al 2021: 12).

It is not just user *data* that is being 'controlled'; compelling and varied cases have been made for the ways in which the commodification of user attention, into standardised attention assets (e.g. impressions, dwell time, clicks) (Hwang *in* Birch et al. 2021: 4), shape the decisions taken by users. Predictions about future user attention and consumption decisions inform the presentation to the user of 'personalised': products, information feeds, advertising, music or videos. A thoughtful or well-judged recommendation based on knowledge of another person is not nefarious in and of itself. Though the technology practices of Web 2.0 have rapidly accelerated the spatial and temporal granularity of these processes and woven them together with business imperatives that have produced monopolistic market power in digital advertising.

Underpinning this are computationally enhanced processes of aggregation and inference that inform prediction based on the construction of taste communities and customer segmentation. 'Behavioural data provides the informational feedback for system control and ultimately profit' (Power 2022: 11) such that deeper questions about the balance of power in the political economy of personal data - and the impacts of this on economies and society - are rightly raised. From this vantage point, digital personal data - in the form of interaction with online services - is measured and classified in ways that construct the metrics and the prediction products, that make these interactions valuable in monetary terms.

Whilst such techcraft is realising value for technology companies from digital personal data - the dimensions of which are defined away from those from whence the data originated - there are also discussions concerning the concurrent erosion of individuals' capacities for self-determination (Zuboff 2019: 519); in part arising from the heavily curated informational ingest already mentioned. Academic contributions from various disciplines have expressed grave concerns over phenomena such as 'cyborgisation' whereby 'reflexivity', the 'defining capability of what it is to be human, deliberative and responsible' (Power 2022: 9), is sacrificed at the alter of a predictive accuracy that directs 'decisions' that have already been made (for us). Some call this a forfeiting of the 'right to the future tense' (Zuboff 2019: 519). I include these concerns in this essay since they have relevance for how we conceive of knowledge building, reality, access and 'value' - all of which may be augmented through new technologies the trajectories of which are not determined. They will be discussed further at the end of the essay.

Given attempts at counterbalancing some of the inequities by the data relations outlined above have been consent mechanisms, its worth briefly mentioning if and where value and information privacy have intersected in the literatures. In the *Privacy as a Right* literatures value is not mentioned explicitly. An individuals' engagement in online commerce activities is considered a loss to them. However, control does feature as a component of privacy (which prompts the question already stated: control over *what*?). By contrast, in the *Privacy as a Commodity* literature the question of the value of personal information comes to the fore, albeit in fairly narrow terms. Studies frame the trading of that personal information, i.e. the control an individual cedes when sharing information, as a 'privacy calculus' where the level of privacy an individual has is conceived as disclosing information or not, and that level is exchanged for a range of perceived benefits at the time of the trade (e.g. personalised services). Given the empirical features of Web 2.0 already discussed (UGC and e-commerce contexts), the theme of personal information and aggregation do feature in this work (Acquisti et al. 2016). Thus the question of the extent to which disclosure of information is contingent on identifiability, or not, is introduced but not speculated about in any great depth. Nonetheless the question of this data as a resource for commercial gain and critical for the development of the digital economy is implicit - which is a notably distinct conception of personal information than in the *Privacy as a Right* literature which deems it as something that can induce harms if adequate safeguards aren't in place.

These literatures acknowledge some (utility) gain on the part of the exchanger of information when interacting with digital services. What has proven far harder to quantify is the extent of those gains, with the question of how to measure the value of free goods a somewhat notorious one. Online choice experiments uncovered that the 'median Facebook user needed a compensation of around $48 to give it up for a month.' (Collis 2020). What is really interesting about this research is how attempts are made to stitch these measures of consumer surplus not currently captured by macroeconomic measures, through the proposal of GDP-B (Brynjolfsson et al. 2019). Thus the reminder that what we decide the value of digital information is, and how we decide it ought to be valued (and what it means to separate the 'what' from the 'how'), has significant normative implications for evaluating growth, direction, progress. By the same token, what is not included in such measures (e.g. the environment, notoriously omitted in measures of GDP), or the costs of harms from social media (how to quantify a loss of self-determination?), have profound implications for our orienteering, or mapping, or measuring, of societies and economies.

III.III.ii. Questions for value in Web 3.0

In Web 3.0 data access control contexts, questions about the value of data may come to the fore since decisions taken about whether to grant access or not may rest less on 'trade-offs' about individually identifiable information, given the different data storage arrangements, and more on understanding what the consequences of the access would be. As understandings of the value of data expand, there could be an emerging set of contingencies that deserve exploration via the question of granting access to personal data based on (as yet) undefined value(s) - where individuals are in closer proximity to those decisions than they are in Web 2.0 contexts. A key question in this context is not only the extent to which individuals may not simply have greater control about whether or not to share 'their' data, but also, greater control over the value defining processes that come to shape the reasons for using the data in the first place.

It culd be that data sharing behaviours facilitated by technologies of Web 3.0 are more suited to theorisation in terms of participation and as such these technologies could be more closely wedded to the work on data institutions, stewards, trusts and co-ops (Aapti Institute 2021; Aapti Institute & Open Data Institute 2021; Manohar 2020a, 2020b; Mozilla Foundation 2020a, 2020b; Mozilla Foundation 2022; Nanda 2021; Ramesh & Kapoor 2020). Thus the question of value will take inspiration from the aforementioned work of civil society groups which stitch people closer to the value defining processes that give their data meaning. Dimensions of control over data may move beyond access and into questions of control over the classification apparatuses that denote salience and legibility in the first place.

	Data storage	Identifiability	Control	Value
Web 2.0	Centralised.	-Low (via PETs). -High (via centralised storage). -Data as an individual medium (DIM REFERENCES). -Data as relational in harms.	Low.	Defined without input from the individuals from whence the data came.
Web 3.0	Decentralised.	-Low (via PETs and decentralised storage). -High (via more granular access controls). -Data as both an individual and relational medium. -Data as relational in value.	High - but control over what exactly?	Defined with input from the individuals from whence the data came?

Table 3. Identifiability, control and value in Web 2.0 and Web 3.0

IV. Between Empiricisms, Between Worlds?

This short essay has outlined some changes underway in a small part of the World Wide Web ecosystem. Namely, a way to end the tight couplings between identity, data and applications. Such changes, should they be adopted, could invert the patterns of digital data accumulation and storage that have characterised Web 2.0. Some effects of this could be that the value generation practices of Web 2.0 are augmented. But it's not clear that all would/should be completely discarded.

This essay has set out some of the literatures that have documented how digital personal data in Web 2.0 contexts is transformed into monetary value. Specifically those that focus on the deeply interwoven processes of classification and prediction and where concerns over the erosion of self-determination and reflexive capacities that result have been raised. This essay has also alluded to the fact that some forms of value generation in this ecosystem evade measurement, and perhaps regardless of data ecosystem context, will continue to do so.

What I've tried to make a case for is that value isn't only a question of what is counted and what is not, or what is attributed a monetary value, or not. But that those decisions are also deeply agentic through their roles in structuring realities. Given this, and the nature of 'personal data' I've suggested that any evaluation of changes in value ought to take identifiability and control, at the level of the individual - but not necessarily (only) in the interest of an individual(s), into account. Well known accounts have discussed the implications of *Seeing Like A State* (Scott 1998) and more recent the implications of *Seeing Like A Market* (Foucade & Healy 2016, also Gandy 1993). If greater powers of legibility were decentralised then what would *Seeing Like A Citizen* produce? Maybe we already know a little about this from how people are choosing to engage with social media. Apparently utopia has not yet been found (which is, perhaps, the point).

Kitchin (2014) called into question the epistemological implications of a Big Data Empiricism which arose alongside Web 2.0 wherein theory could be discarded in place of the truths that would be revealed through the data deluge. As discussed earlier, this impetus became fused with predictive prowess in platform contexts. Moving beyond the requirements of predictive accuracy within the confines of (relatively narrow) commercial imperatives for a moment (e.g. serving the 'right' consumer the 'right' ad), the appeal of predictive accuracy expanded to a form of 'reality mining' in which some academics have in the past been explicit about an 'ultimate goal to create a predictive classifier that can perceive aspects of a user's life more accurately than a human observer (including the actual user)' (Eagle & Pentland 2005: 257). It is precisely the temporally proximate circuitry of the identification of correlation, that inform predictions about consumption behaviour, that executes automated opportunities for commercial activity, that has delivered the astronomical financial successes of technology companies of Web 2.0.

But prediction need not be, indeed is not, the principle objective of data and computational sophistication. Whilst it can be important at the level of policymakers and business imperatives, often (but not always) at the level of an individual it is a foreclosure of possibilities, not an opening (Hong 2022): a compression - not an expansion. When compiling his Paradigms of Science table (**Table 4**), Kitchin sought to draw together some thinking on how Big Data was shifting approaches to science and knowledge building. A fourth paradigm of *Exploratory Science* is said to be underway. Perhaps the evolution of this era could be considered in tandem with the decentralising impetus across Web 3.0 and civil society groups, paying attention to any opportunities for deliberative, but imaginative, decision-making about data, and knowledge, that could open up.

Paradigm	**Nature**	**Form**	**When**
First	Experimental science	Empiricism, describing natural phenomena	pre-Renaissance
Second	Theoretical science	Modelling and generalisation	pre-computers
Third	Computational science	Simulation of complex phenomena	pre-Big Data
Fourth	Exploratory science	Data-intensive, statistical exploration and data mining	Now

Table 4. Paradigms of Science (adapted from Kitchin 2014, compiled from Hey et al. 2009), or: "Whose paradigm is it, anyway?"

Acknowledgements

With many thanks to my supervisor Dr Edgar Whitley for his helpful comments on this paper.

REFERENCES

Aapti Institute. 2021. “Building the Stewardship Navigator—Our approach and methodology”, *Aapti Institute*. https://thedataeconomylab.com/2021/09/28/building-the-stewardship-navigator-our-approach-methodology/.

Aapti Institute, & Open Data Institute. 2021. “Enabling data sharing for social benefit through data trusts”, *Aapti Institute*. https://gpai.ai/projects/data-governance/data-trusts/enabling-data-sharing-for-social-benefit-data-trusts-interim- report.pdf.

Abbasi, A., Sarker, S., Chiang, R. 2016. “Big Data Research in Information Systems: Toward an Inclusive Research Agenda”, *Journal of the Association for Information Systems*, (17:2), pp. i-xxxii https://doi.org/10.17705/1jais.00423.

Acquisti A, Grossklags, J. 2005. “Privacy and rationality in individual decision making”, *IEEE Security Privacy, (3:*1) pp. 26—33.

Acquisti A, John L, Loewenstein, G. 2012. “The impact of relative standards on the propensity to disclose”, *J. Marketing Res.* (49:2) pp. 160-1.

Acquisti, A., Brandimarte, L., & Loewenstein, G. 2015. “Privacy and human behavior in the age of information” *Science*, (347:6221), pp. 509–514. https://doi.org/10.1126/science.aaa1465.

Acquisti, A., Taylor, C., & Wagman, L. 2016. “The Economics of Privacy”, *Journal of Economic Literature*, (54:2), pp. 442–492. https://doi.org/10.1257/jel.54.2.442.

Alessandro, M. 2014. “The future of consumer data protection in the E.U. Re-thinking the “notice and consent” paradigm in the new era of predictive analytics." *Computer Law & Security Review,* 30(6), 643–660. https://doi.org/10.1016/j.clsr.2014.09.004.

Arthur, W. B. 2020. Beijer Institute Discussion Paper 269: Beijer Discussion Paper Series: "Algorithms and the Shift in Modern Science”. https://beijer.kva.se/publication/algorithms-and-the-shift-in-modern-science/.

Bélanger, F., Hiller, J., and Smith, W. J. 2002. “Trustworthiness in Electronic Commerce: The Role of Privacy, Security, and Site Attributes,” *Journal of Strategic Information Systems* (11:3/4), pp. 245-270.

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. 2021. “On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?”. Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 610–623. https://doi.org/10.1145/3442188.3445922

Berners-Lee, T., & Fischetti, M. 1999. “Weaving the Web; The Original Design and Ultimate Destiny of the World Wide Web by Its Inventor” (2 Cassettes). https://dl.acm.org/citation.cfm?id=554764

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.

Birhane, A., & Cummins, F. 2019. “Algorithmic Injustices: Towards a Relational Ethics”. ArXiv:1912.07376 [Cs]. http://arxiv.org/abs/1912.07376.

Boyd, D., & Crawford, K. 2012. “CRITICAL QUESTIONS FOR BIG DATA: Provocations for a cultural, technological, and scholarly phenomenon". *Information, Communication & Society*, 15(5), 662–679. https://doi.org/10.1080/1369118X.2012.678878.

Brynjolfsson, E., Collis, A., Diewert, E., Eggers, F., & Fox, K. 2019. “GDP-B: Accounting for the value of new and free goods in the digital economy”. *NBER Working Paper Series*, Working Paper 25695. http://www.nber.org/papers/w25695.

Carolan, E. 2016. “The continuing problems with online consent under the EU’s emerging data protection principles.” *Computer Law & Security Review*, 32(3), 462–473. https://doi.org/10.1016/j.clsr.2016.02.004.

Collis, A. 2020. “How should we measure the digital economy?”. Submitted to the Department of Management (MIT) on March 3, 2020 in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Management (MIT Sloan School of Management). https://dspace.mit.edu/bitstream/handle/1721.1/126980/1191222397-MIT.pdf?sequence=1&isAllowed=y.

Coyle, D. 2022. “Socializing Data” *Daedalus*, (151:2), 348–359. https://doi.org/10.1162/daed_a_01921.

Coyle, D., & Diepeveen, S. 2021. “Creating and governing social value from data". *SSRN Electronic Journal.* https://doi.org/10.2139/ssrn.3973034.

Coyle, D., & Manley, A. 2021. “Potential social value from data: An application of discrete choice analysis”, *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3973036.

Coyle, D., & Manley, A. 2022. “What is the value of data? Review of empirical methods”, *Policy Brief: Bennett Institute for Public Policy*. https://www.bennettinstitute.cam.ac.uk/wp-content/uploads/2022/07/policy-brief_what-is-the-value-of-data.pdf.

Crawford, K. 2013. “Big data and due process: Towards a framework to address predictive privacy harms”. *New York University School of Law*: Working Paper 13-64.

Dinev, T., McConnell, A. R., & Smith, H. J. 2015. “Research Commentary—Informing Privacy Research Through Information Systems, Psychology, and Behavioral Economics: Thinking Outside the “APCO” Box,” *Information Systems Research*, (26:4), pp. 639–655. https://doi.org/10.1287/isre.2015.0600.

Fourcade, M., & Healy, K. 2016. “Seeing like a market”. *Socio-Economic Review*. https://doi.org/10.1093/ser/mww033.

Gandy, O. H. 1993. “The Panoptic Sort: A Political Economy of Personal Information.” Avalon Publishing.

Ghazawneh, A., & Henfridsson, O. 2015. “A Paradigmatic Analysis of Digital Application Marketplaces”, *Journal of Information Technology*, (30:3), pp 198–208. https://doi.org/10.1057/jit.2015.16.

Gomez Ortega, A., Bourgeois, J., Hutiri, W. T., & Kortuem, G. 2023. “Beyond data transactions: A framework for meaningfully informed data donation”. *AI & SOCIETY*. https://doi.org/10.1007/s00146-023-01755-5.

Hill, G. 2021. “From subject to sovereign”. The Data Economy Lab. https://thedataeconomylab.com/2021/10/22/from-subject-to-sovereign/.

Hong, S.-h.. 2022. “Predictions Without Futures”. *History and Theory*. 61: 371-390. https://doi.org/10.1111/hith.12269.

Inrupt (Producer). 2023a. What is Solid? [video file]. Retrieved from: https://www.inrupt.com/videos/what-is-solid, on: 15.09.23.

Inrupt (Producer). 2023b. What is a WebID? [video file]. Retrieved from: https://www.inrupt.com/videos/what-is-a-webid, on: 15.09.23.

Inrupt (Producer). 2023c. The Recent State of the Web. [video file]. Retrieved from: https://www.inrupt.com/videos/the-recent-state-of-the-web, on: 15.09.23.

Inrupt (Producer). 2023d. What is Web 3.0?. [video file]. Retrieved from: https://www.inrupt.com/videos/what-is-web-3-0, on: 15.09.23.

Kaye, J., Whitley, E. A., Lund, D., Morrison, M., Teare, H., & Melham, K. 2015. “Dynamic consent: a patient interface for twenty-first century research networks." *European Journal of Human Genetics* 23(2), 141–146.

Kitchin, R. 2014. “Big Data, new epistemologies and paradigm shifts”, *Big Data & Society,* (1) pp. 1-12 https://doi.org/10.1177/2053951714528481.

Knijnenburg, B. P., Page, X., Wisniewski, P., Lipford, H. R., & Proferes, N. (Eds). 2022. “Modern Socio-Technical Perspectives on Privacy”. Springer: Open Access ISBN 978-3-030-82786-1 (eBook) https://doi.org/10.1007/978-3-030-82786-1.

Li, Y. 2011. “Empirical Studies on Online Information Privacy Concerns: Literature Review and an Integrative Framework”, *Communications of the Association for Information Systems*, (28:28) pp. 453-496. https://doi.org/10.17705/1CAIS.02828.

Malhotra, N. K., Kim, S. S., & Agarwal, J. 2004. “Internet Users' Information Privacy Concerns (IUIPC): The Construct, the Scale, and a Causal Model," *Information Systems Research* (15:4), pp. 336-355.

Manohar, S. 2020a. “Data Sharing for Public Good: Theoretical Bases and Policy Tools”, *Aapti Institute*. https://thedataeconomylab.com/2020/07/31/data-sharing-for-public-good-theoretical-bases-and-policy-tools/.

Manohar, S. 2020b. “Trust Law, Fiduciaries, and Data Trusts”, *Aapti Institute*. https://thedataeconomylab.com/2020/10/15/trust-law-fiduciaries-and-data-trusts/.

Margulis, S. T. 1977a. "Conceptions of Privacy: Current Status and Next Steps," *Journal of Social Issues* (33:3), pp. 5-21.

Margulis, S. T. 1977b. "Privacy as a Behavioral Phenomenon: Introduction," *Journal of Social Issues* (33:3), pp. 1-4.

Margulis, S. T. 2003a. "On the Status and Contribution of Westin's and Altman' s Theories of Privacy," *Journal of Social Issues* (59:2), pp. 411-429.

Margulis, S. T. 2003b. "Privacy as a Social Issue and Behavioral Concept," *Journal of Social Issues* (59:2), pp. 243-261.

McIntyre, D., Srinivasan, A., Afuah, A., Gawer, A., & Kretschmer, T. 2021. “Multisided Platforms as New Organizational Forms”, *Academy of Management Perspectives*, (35:4), pp. 566–583. https://doi.org/10.5465/amp.2018.0018.

Mozilla Foundation. 2020a. "Shifting Power Through Data Governance”, https://assets.mofoprod.net/network/documents/ShiftingPower.pdf.

Mozilla Foundation. 2020b. “Alternative Data Governance Approaches: Global Landscape Scan and Analysis”, https://assets.mofoprod.net/network/documents/DataGovernanceApproaches.pdf.

Mozilla Foundation. 2022. “Database of Initiatives | Alternative Data Governance in Practice”, Retrieved 18 September 2022, from https://foundation.mozilla.org/en/data-futures-lab/data-for-empowerment/who-is-innovating-database-of-initiatives/.

Nanda, A. 2021. “Policy pathways and governance principles to enable data stewardship”, *Aapti Institute*. https://thedataeconomylab.com/2021/11/18/policy-pathways-and-governance-principles-to-enable-data-stewardship/.

Norberg, P. A., Horne, D. R., & Horne, D. A. 2007. “The Privacy Paradox: Personal Information Disclosure Intentions versus Behaviors,” *The Journal of Consumer Affairs* (41:1), pp. 100-126.

Paré, G., Trudel, M.-C., Jaana, M., & Kitsiou, S. 2015. “Synthesizing information systems knowledge: A typology of literature reviews”. *Information & Management*, (52:2), 183–199. https://doi.org/10.1016/j.im.2014.08.008.

Pentland, A., Lipton, A., & Hardjono, T. 2021. “Building the New Economy: Data As Capital”. MIT Press, Cambridge, Massachusetts.

Ramesh, A., & Kapoor, A. 2020. “Principles for Revenue Models of Data Stewardship”. *Aapti Institute*. https://thedataeconomylab.com/2020/07/31/principles-for-revenue-models-of-data-stewardship/.

Rochet, J.-C., & Tirole, J. 2003. “Platform Competition in Two-Sided Markets”, *Journal of the European Economic Association*, (1:4), pp. 990–1029. https://doi.org/10.1162/154247603322493212.

Scott, J. C. 1998. “Seeing Like A State”. Yale University Press.

Smith, H. J., Milberg, J. S., & Burke, J. S. 1996. "Information Privacy: Measuring Individuals' Concerns About Organizational Practices," *MIS Quarterly* (20:2), pp. 167-196.

Smith, H. J. 2004. "Information Privacy and its Management," *MIS Quarterly Executive* (3:4), pp. 201-213.

Smith, H. J., Dinev, T., & Xu, H. 2011. “Information Privacy Research: An Interdisciplinary Review,” *MIS Quarterly* (35:4), pp. 989–1015. https://doi.org/10.2307/41409970.

Solove, D. J. 2006. "A Taxonomy of Privacy," *University of Pennsylvania Law Review* (154:3), pp. 477-56.

Tene, O. & Polonetsky, J. 2013. “Big Data for All: Privacy and User Control in the Age of Analytics”. 11 *Nw. J. Tech. & Intell. Prop*. 239. https://scholarlycommons.law.northwestern.edu/njtip/vol11/iss5/1.

Thompson, E., & Smith, L. A. 2019. “Escape from model-land". *Economics: The Open-Access, Open-Assessment e-Journal*, 13(1). https://doi.org/10.5018/economics-ejournal.ja.2019-40.

Verhulst, S. G. 2022. "Operationalizing Digital Self Determination”. arXiv. https://doi.org/10.48550/arXiv.2211.08539.

Viljoen, S. 2021. “A Relational Theory of Data Governance”, *The Yale Law Journal*, (82) pp. 573:653.

World Wide Web Consortium (W3). 1998. “A Proposed Convention for Embedding Metadata in HTML”, *Unknown Working Group* - reported by: Stu Weibel. https://www.w3.org/Search/9605-Indexing-Workshop/ReportOutcomes/S6Group2.

World Wide Web Consortium (W3). 2003. “OWL Web Ontology Language - XML Presentation Syntax”, *Web Ontology Working Group*. https://www.w3.org/TR/2003/NOTE-owl-xmlsyntax-20030611/.

Xu, H. 2007. "The Effects of Self-Construal and Perceived Control on Privacy Concerns," in *Proceedings of the 28th International Conference on Information Systems*, Montréal, Canada, December 9-12.

Xu, H., & Dinev, T. 2021. “Reflections on the 2021 Impact Award: Why Privacy Still Matters”, *MIS Quarterly* (46:4), pp. xx-xxxii.

Zuboff, S. 2015. “Big other: surveillance capitalism and the prospects of an information civilization.”, *Journal of Information Technology* 30(1), 75–89.

Zuboff, S. 2019. “The Age of Surveillance Capitalism: The fight for a human future at the new frontier of power”. https://cds.cern.ch/record/2655106.