Workers Collective Data Access Rights

Adding Context to Worker Data Protection

JAKE STEIN*, University of Oxford Department of Computer Science, United Kingdom
DAN CALACCI, MIT Media Lab, USA

How does current data protection law work for workers? The rise of data-driven management practices and workplace surveillance mean that the workplace is increasingly under the purview of data protection laws designed primarily for the consumer context. Workers seeking to use information collected about them as a collective resource—to aggregate, analyze, and share data about their working conditions—may find themselves limited by current data protection and property law. In this paper, we explore how current data protection regimes fall short for worker use and why. We highlight that for workers, data is most valuable when combined with additional context, including potential inferences made about workers that may be excluded from subject access requests. We borrow from both legal scholarship and computer science literature to advance a conception of data access that incorporates notions of mutual legibility and contextual understanding. Drawing on current research into privacy preserving technologies and differential privacy we use formal definitions of privacy from computer science to critique the limited definition of private, identifiable under data protection law. If data access rights are meant to offer a window into how data controllers view subjects, they must also provide access to the context that makes data meaningful in the first place.

CCS Concepts: • **Applied computing** → *Law*.

Additional Key Words and Phrases: privacy, data rights, worker, GDPR, CCPA

ACM Reference Format:
Jake Stein and Dan Calacci. 2022. Workers Collective Data Access Rights: Adding Context to Worker Data Protection. 1, 1 (June 2022), 8 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

The production of value through technology is two-sided. On the one hand is the familiar: value created through selling a new technology as a service or collecting advertising revenue through large streams of user attention. On the other, a new frontier of value creation, presaged by researchers over the past decade, of capturing information on how we do things, and relaying that into value [1]. This model presents a holding pattern which most modern large technology companies have adopted. First, create a product to capture behaviour, create value through insights about that behaviour, and finally use it as raw material for ever-advancing developments in automation and technology, with the goal of manipulating that behaviour in the future. While the immense value created through measuring, hoarding, and analysing consumer behaviour has recently become part of the public, legal, and policy consciousness, this same pattern has been occurring in the world of work for decades, resulting in a process of dual value production in the workplace [2].

Simply by working and going about their daily business, millions of workers around the globe produce data that provides this same raw material. As in the consumer context, these new patterns of value creation and extraction have transformed everyday workplace behaviours into a valuable resource, used to make decisions or train algorithms that can have profound impacts on hiring practices, wages, working conditions, and workers' ability to organise and build power. Hiring firms like HireVue score job applicants based on interaction data captured from video interviews, while people management software like WorkDay allow single-pane-of-glass views into worker productivity, and pay for management [3]. An increasing group of tools transform everyday office communication into performance monitoring metrics used for reviews and even firing decisions [4]. Detailed data about warehouse workers is used to set production quotas and work requirements that can result in injury or workplace labour violations [5], and careful monitoring of employees is used to predict unionisation likelihood [6]. Online platforms like Shipt, DoorDash, and Instacart leverage opaque algorithmic pay systems that use worker data to optimise worker wages [7].

As firms move increasingly towards leveraging algorithmic decision-making in the pursuit of efficiency, this data collection has become more intrusive, with some systems tracking employees' movements and activities on and off the clock [8–10], prompting some scholars to make fundamental privacy arguments about the raw data collected. And as [9] notes, while the data collected by an employer may seem innocuous, such as through a workplace wellness program or social media tracking in hiring, the inferences capable from such data, especially in combination with other datasets, can include information as private as as sexual orientation, personality traits, and medical status.

All of these advancements in the world of data at work fundamentally mean that there has been a clear shift in power between the worker and the employer in the digital age. Although employers have always had the prerogative to surveil employees to some degree under common law, the radical change in degree of surveillance and range of possible inferences employers can now make about workers mean that the game has fundamentally changed. In this paper, we investigate the options available to workers who wish to change its rules through collecting, aggregating, requesting, or otherwise using the data they produce at work for collective sense-making, organising, and auditing of employer practices; actions not

[*] Both authors contributed equally to this research.

Authors' addresses: Jake Stein, jake.stein@cs.ox.ac.uk, University of Oxford Department of Computer Science, Oxford, United Kingdom; Dan Calacci, MIT Media Lab, , Cambridge, USA, larst@affiliation.org.

XXXX-XXXX/2022/6-ART $15.00
https://doi.org/XXXXXXX.XXXXXXX

only crucial to maintaining a healthy economy, [11] but also for worker well-being and fundamental dignity.

Existing law treats any data an employee produces at work as prima facie the employer's property in many jurisdictions [12], severely limiting the ability of employees to access or use the data they produce at work. But recent data privacy laws such as the GDPR in the EU and the CCPA in the US may offer avenues to workers who wish to access and interrogate the personal data used to manage and define their employment relationships.

While the general question of data rights at work has been addressed in discussions elsewhere [6; 12; 13], in this paper we investigate the current boundaries imposed on workers who seek to collect, request, aggregate, and otherwise use the data they produce while at work for collective sense-making and organising. We first provide an overview of the kinds of data workers create while at work, drawing from research on workplace surveillance to taxonomize the information workers produce while in the employment relationship. We then investigate the avenues that a modern worker has for accessing this data, either through data access requests or other means, and the implications of current law in varied jurisdictions on how such data may be used.

1.1 What's at Stake: Worker Data Today

Access to information is core to the way that modern firms control workers and how worker groups organise. Over the past decade, the steady spread of workplace monitoring technology has turned places of work into places where workers are quantified and monitored constantly [14]. While the data collected from workers has different uses depending on the work context, it invariably flows unilaterally from the worker to the employer, and is used most often to control and manage workers.

Nowhere is this more visible than in the case of platform work. Fissured workers for firms like Uber or Lyft are controlled almost exclusively through what is referred to as information asymmetry. This refers to the ways that platforms and algorithms strategically limit the information made available to workers in order to incentivize certain behaviors [15]. For example, by limiting information about jobs made available to drivers, platforms effectively enforce limited choice in what orders workers can accept. These strategies are part of the "soft control" that workplaces are instituting using data collected from workers generally.

Using data to monitor and control workers is far from limited to the platform context. The majority of US companies monitor their employees' internet use or log their workers' keystrokes [14; 16–18]. Meanwhile, the coronavirus pandemic has expanded the surveillance prerogative enjoyed by employers into the homes of workers who are working from home. Workers are increasingly required to install software on their computers and phones that track their location, screenshot their computers, and even record their working environment through their webcams, sometimes without workers' knowledge [12; 19; 20].

As scholars Niels van Doorn and Adam Badger convincingly argue, the data that workers produce at work also has real value. They argue that firms knowingly engage in a kind of "dual value production", wherein they profit both from the functional use of data collected from workers and the "speculative value" of data as assets to be sold or traded [2]. The limited reach of worker's access to this data also severely limits their ability to share in the value created from their labor.

The information workers produce also has potential as a useful organizing and power-building tool for workers, but only if they are able to access and wield it as a collective right. The history of worker response to scientific management demonstrates that even early responses by organized labor to new management technologies involved self-directed data collection, or negotiating access to the information that employers hoarded about workers [21; 22]. Information sharing generally also has deep roots in the study of labor organizing. Rather than only generating power through withholding labor, a major way that organized workers build political power is through leveraging information and "voice" to understand how worker and employer interests intersect and relate [23; 24]. Bargaining models of labor economics also demonstrate that access to information can fundamentally change the political trajectory of worker movements: access to information enables groups to fight for the "median" worker, rather than the "least attached" member of a worker group [6; 24].

2 LEGAL BASIS AND STUMBLING BLOCKS: PRECEDENTS FOR COLLECTIVE WORKER DATA ACCESS

The main tools available to workers to aggregate data collected about them at work rests in data protection law. A consistent aim across data protection law has been to empower data subjects with access to personal data held by controllers [25; 26]. This entitlement would appear to aim for data subjects to, normatively speaking, know what controllers know about them. Despite apparent consensus that data subjects should have some right to data access, the specific objectives and remit of such rights remain opaque and tangled with conflicting rights of other stakeholders, further complicating the question of how workers might gain meaningful access over their data [27].

In this section, we summarise the legal basis on which subject data access can be pursued through data protection law. With this in mind, we move on to highlight how legal definitions of data as simultaneously the object of personal privacy for subjects and of commercial property for controllers create a paradox which shuts workers out from sufficient access to information or use of their data [6; 12; 14].

2.1 Data Protection and Access Rights

The legislation that defines rights for data access reveals dividing lines with respect to control over personal data and the extent of control given. Interestingly, the EU's GDPR does not directly refer to privacy as the goal of data access rights, instead referencing the fundamental right to, "data protection" established in its founding charter [28]. GDPR's Article 15 further entitles data subjects to "the right to obtain from the controller confirmation as to whether or not personal data concerning him or her are being processed," along with supplementary rights to: "any available information as to [the data's] source…the existence of automated decision-making, including profiling…meaningful information about the logic involved,...the significance and the envisaged consequences of such processing for the data subject" and finally, "…a copy of the personal data" [29]. Similarly, California's 2018 Consumer Protection Act, requires businesses to disclose "the categories and specific pieces of personal information" and rights to correct or delete personal data as well as a copy of the data itself [30]. While these data protection laws may initially seem sufficient for supplying individual workers with strong claims not only to their personal data, but also to much of the relevant contextualising information about data's use, each contains critical caveats preventing workers from applying the rights in practice to achieve meaningful access.

The CCPA is fairly straightforward in how it limits workers' ability to take advantage of these rights [30]. Section 1798.145 explicitly excludes employees and contractors from data access rights, instead, only provisioning them the right to notification of the data which will be collected about them.

The GDPR is more generous, extending the same data access request rights to employees as it does to consumers – so long as requesters meet the definition of data subject. Further, legal precedent in the ECHR, has previously found that data produced through work systems or while at work can be considered the personal data of employees even preceding GDPR's adoption [31].

However; GDPR also arms controllers with further tools to avoid providing workers with access to data. Data access must not interfere with other rights, "including trade secrets or intellectual property and in particular the copyright protecting the software" [29]. For workers seeking to better understand their own activities, the algorithms that govern their pay or dismissal, or even to contextualise their personal data amongst their peers, this statement can preempt their claim to their own data.

Data's value is dependent on its context. Constraining access to rich, contextualised data sets for workers, in practice, removes their ability to access information about themselves as individuals too [32]. The dual legal definition of data as subject to data protection rights for individuals, but simultaneously the private property of controllers, leaves workers in a situation where they would appear to have access to copies of their own personal data, but in practice are prevented from aggregating or pursuing the information contained in their data as a collective set. The paradox established by the current legislative climate allows workers a claim to access their personal data, but by disallowing legibility within collective data in practice, places the information personal data contains out of workers' reach.

2.2 Trade Secrets and Intellectual Property

Arguably the largest barrier to workers' ability to make collective sense of the data they produce at work may be trade secret, intellectual property, and copyright law. The amorphous nature of trade secret application means that a wide variety of data—even information that is readily ascertained by employees—may be legally protected as employer property. In particular, trade secret laws are able to protect information and inferences made collectively, even if individual-level data is not protected.

For example, while truck drivers may be able to individually collect data on the routes they take in order to e.g. track mileage, the aggregation of a fleet of truck drivers' routes may encompass a trade secret insofar as it might reveal the locations an employer services, a well-established area of trade secret law. Aggregate collection of wage or salary data has also been the aim of trade secret litigation, although the strength of the trade secret argument to salary transparency has been questioned by legal scholars [33].

However, trade secret is not representative, as Cynthia Estlund notes, of a broad right to secrecy [34]. Existing labour rights conferred to US workers by the National Labor Relations Act are clear about workers' ability to discuss wages and work conditions with each other and potential public allies. Trade secret law is intended to protect firms from private misappropriation of information, not to hinder worker power. As Estlund convincingly argues, employers should be limited in trade secret claims when it comes to information already disclosed to workers that has legitimate value in public disclosure [34].

Notably, none of this reasoning offers a definitive answer to whether or not inferences made by employers about employees should be considered exempt from trade secret, or should be covered under acts such as the NLRA. Such inferences may also rest under intellectual property and copyright laws instead. Inferences about employees, which we argue could be considered personal data, then demonstrates a conflict between privacy claims and notions of data as private property.

2.3 The "Data Protection Conundrum": Workers' Private Data, Employers' Private Property

Two 2021 cases brought by workers of the App Drivers and Couriers Union (ADCU) resulted in drivers receiving significant swathes of their data previously withheld by platforms [35; 36]. Ride hailing platforms Ola and Uber claimed that the data and the algorithmic systems produced through use of that data were protected under trade secrets and intellectual property law [37]. Though the ADCU was able to achieve visibility into withheld data from Uber, it was barred from accessing data in a format conducive to computing or aggregation. The split finding of the case perfectly illustrates the two elements of the data access paradox. Sofaras users want individual access to their personal data, it is very much their right as was upheld by the court, but should workers want to aggregate data, data controllers are entitled to obstruct those efforts. Accordingly, data can exist both as personal for workers, while the knowledge derived from that data is guarded as the property of controllers. The result is a legal environment where data subjects have access to their personal data, but lack meaningful access, in this case the "ability to validate the fare basis and compare earnings and operating costs" [36; 37].

Of further interest is the dispute leveled against the use of data for re-aggregation. Applicants of the Ola case requested their data so that it may be aggregated in a data trust to strengthen the position of app drivers and their union. Ola, in response argued: "The request to transfer personal data in a certain format stems from the wish of [applicants] to have this data entered directly in a WIE database for analysis with the aim of improving the negotiating position of platform workers. Recital 68 of the GDPR states that the right to data portability serves to strengthen the data subject's control over their own data" [35]. The difference Ola sees between those statements is intent to benefit a collective of workers, not any individual worker. Allowing re-use in the interest of algorithmic transparency "only serves the interests of ADCU or the general interests of drivers who use platform services." The distinction Ola attempts to make is between individual portability on the basis of personal data protection and the collective interest served by the aggregation of data: "the [applicants] use the right of access and the right to data portability for a purpose other than that for which it was given, namely to set up a data trust and to gather information to improve the legal position of drivers". From Ola's point of view, data in its aggregate context is an asset protected under intellectual property per GDPR's specification that "the right to obtain a copy referred to in paragraph 3 shall not adversely affect the rights and freedoms of others," particularly then it strengthens the legal positions of workers. [36].

Though the court ruled in favour of applicants' right to re-use personal data in an aggregate setting, it also reaffirmed that the basis for data portability is a fundamentally economic one – and one defined in the terms of software providers. Data portability, Ola argued, is designed for anti-competition and portability for users to move seamlessly between platforms with similar aims, but not necessarily for the redistribution of information for use by worker (or consumer) collectives and unions. The relevance of such information to the relationship between controller and subject or the need for contextualising individual data via re-aggregation, would appear (to Ola) to be outside the scope of use for DSARs. In this sense, DSARs' basis in individual data protection is beneficial to controllers, allowing them to fein access for subjects, while claiming exclusive control over the value of data in its aggregate form. Such a combination – the paradoxical perspective on data rights from personal data protection and property viewpoints – reifies the subordination of subjects and their rights to data (worker and consumer alike) vis-a-vis data controllers.

This ruling creates precedent for collectives to apply DSARs to data controllers in the interest of aggregating and analysing data in a collective context. However, the ruling also holds the door open for claims of intellectual property violations to be levied against collectives by affirming data portability rights are derived with the purpose of movement between platforms who are conveniently unperturbed by allowing for individual privacy and access, but work tirelessly to prevent users from accessing aggregating data in the collective context. Bodie offers a solution of shared governance between workers and controllers, with controllers acting as information fiduciaries, even referencing UK app driver actions as a particularly strong example of data portability for the sake of aggregation [12]. However, exemplified in these two cases, it has become clear that companies are tending in precisely the opposite direction of such shared governance, not only hoping to keep data closed down, but using a privacy-oriented interpretation of data subject access rights as a means to justify obstructing subject access.

Newman also points out the paradox of individual privacy rights and property rights being applied to the same data in its disaggregated and aggregate forms. Accordingly, he further highlights that employers point to the individual privacy of workers as a means to preempt claims over collective data. However, Newman gives a positive example of one US union that was able to win access to individual data in its collective form, but only when it met anonymization criteria; "the need to enforce collective rights overrode attempts by the employer to assert that imagined individual-privacy rights of other employees." It's also important to note workers' status as employees, contractors, or merely users of apps for work is increasingly embattled, inevitably affecting the applicability of data access rights. California's proposition 22, which allowed app drivers to continue to be classified as contractors and was the subject of a multi-billion dollar lobbying and occasionally illegal worker harassment campaign by Uber and Lyft meant that drivers in the state (as contractors) don't retain the rights of consumer data access under the CCPA, but simultaneously cannot benefit from full-time employee status and associated benefits of labour organisation [38].

A few convergent factors within data protection law create a paradox in which workers might claim information as their personal data, but not the right to access that same data meaningfully due to the protection of information created in aggregate under intellectual property and trade secrets protections specifically exempt in data protection law.

First, data protection fails to define the end purpose of data access rights, specifically, though counterintuitively avoiding privacy. Second, workers' status as employees or consumers with respect to controllers is continually made ambiguous, making the level of access they should be afforded ambiguous with it [12]. Finally, workers are often explicitly excluded from rights to data access, affirming precedence of employer property claims or privacy claims of workers.

3 WHAT WORKS FOR WORKERS?

Thus far, we have explored the landscape of worker data, its potential value for workers, and a legal climate which undercuts workers' ability to realise sufficient data access to understand their own activities at work to a similar depth as their employers. We argue that mutual legibility – access to the full richness and potential that collective data contains – is a crucial component of personal data access. In the following section, we sketch out the potential strategies to address existing asymmetries that have been proposed in both legal and computer science literature to make mutual legibility a more realistic target. Further, we identify research in privacy preserving data analysis techniques and intermediaries as an undervalued source of formally defined standards for access to collective data by workers.

A broad variety of approaches have been proposed to enhance or extend transparency into subject data and algorithmic systems. However, enabling more comprehensive data access for workers touches on such a broad variety of ongoing debates in both computer science and the law that no one element proposes a universally applicable framework. We class the most relevant solutions that might contribute to greater data access into a few categories.

3.1 Expanding Definitions of Personal Data and the Remit of Privacy Rights

One solution may be to expand definitions of personal data to include additional contextual information or even data products derived from "raw" personal information. Scholars from both the technological and legal perspectives have proposed solutions that seek to incorporate greater levels of explainability for algorithmic decision-making into data protection [33; 39–41]. Most notably, Wachter and Mittlestadt expand data protection rights to include a "right to reasonable inference," compelling controllers to report;

> (1) why certain data form a normatively acceptable basis from which to draw inferences;
> (2) why these inferences are relevant and normatively acceptable for the chosen processing purpose or type of automated decision; and
> (3) whether the data and methods used to draw the inferences are accurate and statistically reliable [42].

In the context of worker data access, rights to explanations would certainly advance transparency into algorithmic systems driven by worker data. However, this right alone remains focused on data protection and access on the level of the individual worker, contextualising independent decisions, without empowering workers with the full purpose of accessing their data – understanding their data, as their employer does. Further, these rights compel controllers to define and disclose relevant patterns and inferences in the data, but could very well miss patterns implicitly included in decisions, which might be easily caught with the context workers possess. Also, the right is geared toward ameliorating decisions for individual data subjects, when decisions are made (especially algorithmically) though the use of aggregate data indirectly drawing on innumerable other data subjects' personal data to define patterns or correlation, patterns that should be visible to collectives too. Finally, the level of explainability often argued for might not always be technically feasible among massive data sets or black-box algorithmic systems, especially when compared to informed deductive data analysis on underlying data sets used to train models.

Expanding data protection rights to inferences and explainability would represent an enormous gain for workers, but the gaps it leaves for worker collective access also point to a need for greater access to underlying data to improve transparency of the collective knowledge contained in datasets beyond individual inferences. A more bottom-up approach to data collection could potentially fill such a gap.

3.2 Enabling Collective Self-Inquiry

One other outlook is to create policy that can support a "bottom-up" approach to collecting, aggregating, and analysing worker data by workers themselves [43–45]. Drawing on empirical work charting how Deliveroo gig workers use a patchwork of supplementary technologies to coordinate resistance against the app's algorithmic management mechanisms to optimise pay and better working conditions from the perspective of workers, Woodcock coined the term "digital workers' inquiry" [46; 47]. Inquiry prioritises "first, that we need to understand how technology is experienced by workers in the labour process; [and] second, to understand how technologies are being appropriated for workers' use" [46]. Such a bottom-up approach has immediate appeal [43]. Workers need not depend on the controller to provide data or even collect it in order to amass the necessary information to immediately enhance their autonomy. Data could be collected and aggregated in a format that respects stakeholders preferences and needs while easing the process of finding information relevant to workers' specific concerns [27]. If done independently, inquiry could also sidestep controllers' complaints of the burden of responding to DSARs [48].

Examples of self-inquiry take a variety of forms. WeClock and Driverseat rely on purpose-built applications that allow workers to log their movements during the day, collecting data that is later aggregated to provide contextualised insight into working conditions [49; 50]. On the other hand, approaches like those of the Worker Info Exchange (discussed earlier for their lawsuits) go about self-inquiry by way of collecting DSARs in a data trust for further analysis rather than via self-tracking on personal devices [51].

While self-inquiry provides options to fill gaps left by arguments to expand data protection to inferences, it also comes with risks of its own. First, it is only effective in concert with the abovementioned rights as it provides no guarantees that data collected would be relevant to patterns of discrimination, or instances of injustce in workers' expereinces. Whereas expanding rights to inferences overfits to individual subjects and contested decisions, self-inquiry might under-fit, providing a wealth of data, but no guaranteed information for a relatively high bar of effort for workers. Further, controllers will inevitably have access to data sets they might combine with worker data to enrich its meaning that workers cannot collect themselves (e.g. market research, customer data, systems telemetry data). Also, the "fissuring" of workers, the institutional clumsiness of older union organisations, and slow development of bottom-up data trusts makes developing sophisticated enough data analytics infrastructures to rival controllers' armies venture-backed silicon valley engineering cores a daunting task [52; 53]. Finally, given the strength of property claims to data collected at work, it's not hard to imagine intellectual property and trade secrets claims being levied against data amassed in efforts of collective self inquiry, effectively creating worker-led expansion of workplace surveillance [6].

3.3 Mutual Legibility and Data's Determination

Some scholars offer more expansive and fundamental interpretations of how asymmetric data aggregation, particularly in the collective case, interferes with data access rights. Delacroix and Veale, for instance, argue that access to one's data in its full context concerns a matter of freedom of self-determination and identity above and beyond privacy and data protection [54].

For the authors, the shortest route to autonomy might not be found in the additional data collection or "counter-profiling" that we might associate with self-inquiry or, "addressing epistemic imbalances (e.g. through 'explainable' systems)" we could ascribe to expanded rights to transparency. Instead, they offer it might be in making access to elements of aggregate data for both controllers and to data subjects more opaque [54]. This approach is certainly applicable in the case of worker data, and introduces a critical tool for data protection: obfuscation. Perhaps enforcing controllers and workers to both have less behavioural data is an effective alternative towards mutual legibility. Lee et al. illustrate one way this might be possible, applying participatory design methods involving employers and workers together to reorient how workplace data collection is implemented and the governance that controls it [55].

That is not to say this approach is mutually exclusive to the two routes already described. To the contrary, recent research in the field of privacy preserving technologies and anonymization techniques illustrate a way to operationalize these concepts together.

3.4 Computational Privacy Research as a Middle Path

Balancing the same tradeoff between obscuring features of a data set [56; 57], or exposing them to the right stakeholders [58] is at the centre of discourse on privacy preserving technologies and architectures, but is rarely brought into questions of data access rights [59]. Contrary to legal approaches advocating for the expansion of rights, which depend on algorithmic explainability that is not yet technically feasible, or the technological approach of self-inquiry, which likewise depends on legal interpretations which may not persist, looking to privacy research and specifically, “functional anonymization” grants a more robust template for determining what data workers should have access to agnostic of legal or technical implementation [60].

Paul Ohm resolutely asserts that, "data can be useful or perfectly anonymous but not both" [57]. Mapped onto the context of worker data access, this concept can be interpreted as; data that is not anonymous (identifiable), should be personal data, and thus available to subjects, but when anonymous or individualised enough to avoid conflicting property claims from controllers, data is no longer meaningful for subjects or controllers alike. Ohm’s dilemma resonates perfectly with the paradox of individual worker data access vs. meaningful collective data access. Recall Uber’s argument that the data requested by workers might violate the individual privacy of customers should it be disclosed to the degree necessary for it to be useful to ADCU, or Newman’s case in which postal worker data would only be disclosed to the unions if it was stripped of demographic characteristics [6; 35].

Elliot et al. push back against Ohm’s binary, offering a way out via a form of anonymization that also involves a degree of disclosure, in a similar, but inverted perspective to how Delacroix and Veale argue for gains in agency and autonomy via obfuscation [54; 57; 60]. Functional anonymization is defined as the principle that “whether data are anonymous or not (and therefore personal or not) is a function of the relationship between those data and their environment” [60]. The most important element of the environment being “any data the set might be joined with, the context in which the data could be used and thus re-identified,” as well as the users, governance, and institutional infrastructures in which the data at question is situated [60].

We are not arguing that controllers adopt functional anonymization and consider the job done. To the contrary, we posit that the definition of an extended field of what constitutes private data through the environment that functional anonymization spotlights should also be disclosed to collectives as a form of personal data that becomes private as a result of the collective context or aggregation.

Functional anonymization lends an effective framework for defining what in the existing literature is a mystery – how a right to data access (particularly when motivated by individual privacy) corresponds to the knowledge or information that can be derived from that information only in its collective form. There are already proposals for how to put this interpretation into practice. Binns and Veale for instance, propose data intermediary architectures which balance the proper volume of disclosure to prevent algorithmic discrimination, while also preserving the privacy of individual users [58]. One might imagine that the features of data included as a result of this extension could be made available via an intermediary when they reach a threshold of collective relevance.

Of course this approach also has its weaknesses. The most realistic means for the implementation of this interpretation of data access – data intermediaries like those proposed by Binns and Veale – are not yet mature enough and still lack sufficient means to convince controllers to allocate data, though they may be more compelled to do so given the expanded interpretation of data access might resolve some of the conflicts with rights cited by controllers. Purtova in particular notes that using principles of anonymization and identification that dilute personal data to include elements like those we propose here also poses serious danger to data protections’ potency moving forward, when “everything is information” [61]. On the contrary; the approach of using anonymization as a means to identify what should be accessible to collectives, much like the concept of functional anonymization on which it is based, should be understood as an approach of incrementalism with respect to a status quo in which little to no collective access exists at all.

4 CONCLUSIONS AND RECOMMENDATIONS

We have highlighted the contradictions and paradoxes at the centre of worker data access in practice, moving the discourse past theoretical ruminations on how to achieve meaningful data access for collectives to a focus on tools and obstacles in practice. Further, we have presented the miscellany of tools workers have at their disposal for advancing their understanding of themselves in context.

Current regimes define data in two main ways: the object of privacy, and the object of property. These two definitions are in many ways fundamentally opposed, but often apply to the same data simultaneously. When examined in the context of worker-created or related information, where value is generated primarily through collective action and aggregation, the tension between these two frames becomes even clearer. For data to be used meaningfully in the labor context, laws regarding its protection and use need to incorporate some conception of data as relating to people, contexts, and ideas of legibility.

Considering data in this way, it becomes clear that current data protection and rights frameworks—which operate at the individual level—are severely limiting. They cannot provide full legibility, even to individuals, as most data collected and governed by data protection is made meaningful only through additional context that is either beyond the scope of current regimes or limited by property law.

How can data protection law address this? We may be able to develop technical or legal definitions of meaningfulness around data, such as following the line of reasoning that functional anonymity provides. These definitions might be able to guide what additional information should be provided to data subjects, but as of now, such solutions are immature and difficult to generalise. Instead, perhaps the contextual nature of the legibility of data should be addressed by contextual laws. For example, California's AB-701 allows warehouse workers to request and aggregate data about their working conditions—working quotas and work speed data, among other data points—if they suspect they have been subject to labor violations. Laws like AB-701 turn data access rights into rights available to people in specific contexts, with specific limits and aims. To be sure, AB-701 cannot address all of the potential information that warehouse workers and organisers may be interested in having access to, but it is at the very least a significant step forward for worker data rights.

Although data protection and privacy regimes fundamentally address notions of individual agency and identity, they fail to address the degree to which such ideas depend on contextual understanding. By treating individual data as property or items people have extended rights to—rather than objects that have meaning mainly within a larger context—they strip data subjects of some fundamental rights to self-legibility.

ACKNOWLEDGMENTS

REFERENCES

[1] McKenzie Wark. *Capital Is Dead: Is This Something Worse?* Verso, 2021.

[2] Niels van Doorn and Adam Badger. Platform Capitalism's Hidden Abode: Producing Data Assets in the Gig Economy. *Antipode*, 52(5):1475–1495, 2020.

[3] Maria Aspan. A.I. is transforming the job interview—and everything after. *Fortune*, (Special Report on Artificial Intelligence), January 2020.

[4] Esther Kaplan. The Spy Who Fired Me: How everything became trauma. *Harper's Magazine*, March 2015, March 2015.

[5] Jodi Kantor, Karen Weise, and Grace Ashford. The Amazon That Customers Don't See. *The New York Times*, June 2021.

[6] Nathan Newman. Reengineering workplace bargaining: How big data drives lower wages and how reframing labor law can restore information equality in the workplace. *U. Cin. L. Rev.*, 85:693, 2017.

[7] Bryan Menegus. 'Every Single Person Is Losing Money': Shipt Is the Latest Gig Platform to Screw Its Workers. https://gizmodo.com/targets-shipt-pay-model-change-cuts-worker-pay-shipter-1841620656, February 2020.

[8] Kaveh Waddell. Why Bosses Can Track Their Employees 24/7. https://www.theatlantic.com/technology/archive/2017/01/employer-gps-tracking/512294/, January 2017.

[9] Pauline T. Kim. Data Mining and the Challenges of Protecting Employee Privacy under U.S. Law. *Comparative Labor Law & Policy Journal*, 40(3):405–420, 2018.

[10] Lothar Determann and Robert Sprague. Intrusive monitoring: Employee privacy expectations are reasonable in Europe, destroyed in the United States. *Berkeley Tech. LJ*, 26:979, 2011.

[11] Arne L. Kalleberg, Michael Wallace, and Robert P. Althauser. Economic segmentation, worker power, and income inequality. *American journal of Sociology*, 87(3):651–683, 1981.

[12] Matthew T. Bodie. The Law of Employee Data: Privacy, Property, Governance. *Indiana Law Journal*, 97, 2021.

[13] Jeremias Adams-Prassl. What if your boss was an algorithm? Economic Incentives, Legal Challenges, and the Rise of Artificial Intelligence at Work. *Comp. Lab. L. & Pol'y J.*, 41:123, 2019.

[14] Ifeoma Ajunwa, Kate Crawford, and Jason Schultz. Limitless worker surveillance. *Calif. L. Rev.*, 105:735, 2017.

[15] Alex Rosenblat and Luke Stark. Algorithmic labor and information asymmetries: A case study of Uber's drivers. *International Journal of Communication*, 10:27, 2016.

[16] Kirstie Ball. Workplace surveillance: An overview. *Labor History*, 51(1):87–106, 2010.

[17] Antonio Aloisi and Elena Gramano. Artificial Intelligence Is Watching You at Work: Digital Surveillance, Employee Monitoring, and Regulatory Issues in the EU Context Automation, Artificial Intelligence, & Labor Law. *Comparative Labor Law & Policy Journal*, 41(1):95–122, 2019.

[18] Olivia Solon. Big Brother isn't just watching: Workplace surveillance can track your every move. *The Guardian*, November 2017.

[19] Bobby Allyn. Your Boss Is Watching You: Work-From-Home Boom Leads To More Surveillance. *NPR*, May 2020.

[20] Adam Satariano. How My Boss Monitors Me While I Work From Home. *The New York Times*, May 2020.

[21] Vera Khovanskaya and Phoebe Sengers. Data Rhetoric and Uneasy Alliances: Data Advocacy in US Labor History. In *Proceedings of the 2019 on Designing Interactive Systems Conference*, pages 1391–1403, San Diego CA USA, June 2019. ACM.

[22] Vera Khovanskaya, Lynn Dombrowski, Jeffrey Rzeszotarski, and Phoebe Sengers. The Tools of Management: Adapting Historical Union Tactics to Platform-Mediated Labor. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW):1–22, 2019.

[23] Albert O. Hirschman. *Exit, Voice, and Loyalty: Responses to Decline in Firms, Organizations, and States*, volume 25. Harvard university press, 1970.

[24] Richard B. Freeman and James L. Medoff. What do unions do. *Indus. & Lab. Rel. Rev.*, 38:244, 1984.

[25] René L. P. Mahieu, Hadi Asghari, and Michel van Eeten. Collectively exercising the right of access: Individual effort, societal effect. *Internet Policy Review*, 7(3), July 2018.

[26] European Parliament and of the Council of 27 April 2016. EU General Data Protection Regulation (GDPR), April 2016.

[27] Jef Ausloos, René Mahieu, and Michael Veale. Getting Data Subject Rights Right, November 2019.

[28] Charter of Fundemental Rights of the European Union, December 2000.

[29] General Data Protection Regulation Chapter III, 2016.

[30] California Consumer Privacy Act, June 2018.

[31] Copland vs. United Kingdom, April 2007.

[32] Durant v Financial Services Authority, December 2003.

[33] Cynthia Estlund. Extending the case for workplace transparency to information about pay. *UC Irvine L. Rev.*, 4:781, 2014.

[34] Cynthia Estlund. Just the facts: The case for workplace transparency. *Stan. L. Rev.*, 63:351, 2010.

[35] Applicants 1-10 vs. UBER B.V., November 2021.

[36] Applicants 1-3 vs. OLA NETHERLANDS B.V., November 2021.

[37] Natasha Lomas. Dutch court rejects Uber drivers' 'robo-firing' charge but tells Ola to explain algo-deductions, December 2021.

[38] Megan Rose Dickey. Uber drivers sue company alleging coercive Prop 22 advertising, October 2020.

[39] Isabel Ebert, Isabelle Wildhaber, and Jeremias Adams-Prassl. Big Data in the workplace: Privacy Due Diligence as a human rights-based approach to employee privacy protection. *Big Data & Society*, 8(1):20539517211013051, 2021.

[40] Joshua A. Kroll. Outlining Traceability: A Principle for Operationalizing Accountability in Computing Systems. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, pages 758–771, Virtual Event Canada, March 2021. ACM.

[41] M. Pégny, E. Thelisson, and I. Ibnouhsein. The Right to an Explanation. *Delphi - Interdisciplinary Review of Emerging Technologies*, 2(4):161–166, 2019.

[42] Sandra Wachter and Brent Mittelstadt. A right to reasonable inferences: Rethinking data protection law in the age of big data and AI. *Colum. Bus. L. Rev.*, page 494, 2019.

[43] Sylvie Delacroix and Neil D Lawrence. Bottom-up data Trusts: Disturbing the 'one size fits all' approach to data governance. *International Data Privacy Law*, page ipz014, October 2019.

[44] Nick Couldry and Alison Powell. Big data from the bottom up. *Big Data & Society*, 1(2):2053951714539277, 2014.

[45] Jack Hardinges and Jared Robert Keller. What are 'bottom-up' data institutions and how do they empower people? – The ODI, June 2021.

[46] Jamie Woodcock. Towards a Digital Workerism: Workers' Inquiry, Methods, and Technologies. *NanoEthics*, 15(1):87–98, 2021.

[47] Sai Englert, Jamie Woodcock, and Callum Cant. Digital Workerism: Technology, Platforms, and the Circulation of Workers' Struggles. *tripleC: Communication, Capitalism & Critique. Open Access Journal for a Global Sustainable Information Society*, 18(1):132–145, January 2020.

[48] Laura Diaz Silvarrey. Response to DCMS consultation "Data: A new direction". page 89, 2021.

[49] WeClock.it. https://weclock.it/, 2021.

[50] Driver's Seat Cooperative. https://driversseat.co/, 2020.

[51] Worker Info Exchange | Data Rights for Digital Workers | London. https://www.workerinfoexchange.org, 2021.

[52] David Weil. *The Fissured Workplace: Why Work Became So Bad for So Many and What Can Be Done to Improve It.* Harvard University Press, February 2014.

[53] Aditya Singh and Jack Hardinges. Measuring the Impact of Data Institutions. Technical report, Open Data Institute, London, March 2022.

[54] Sylvie Delacroix and Michael Veale. Smart technologies and our sense of self: Going beyond epistemic counter-profiling. In *Life and the Law in the Era of*

Data-Driven Agency. Edward Elgar Publishing, 2020.

[55] Min Kyung Lee, Ishan Nigam, Angie Zhang, Joel Afriyie, Zhizhen Qin, and Sicun Gao. Participatory algorithmic management: Elicitation methods for worker well-being models. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, pages 715–726, 2021.

[56] Cynthia Dwork. Differential Privacy. In Michele Bugliesi, Bart Preneel, Vladimiro Sassone, and Ingo Wegener, editors, *Automata, Languages and Programming*, pages 1–12, Berlin, Heidelberg, 2006. Springer Berlin Heidelberg.

[57] Paul Ohm. Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization. *UCLA l. Rev.*, 57(6):1701–1778, 2010.

[58] Michael Veale and Reuben Binns. Fairer machine learning in the real world: Mitigating discrimination without collecting sensitive data. *Big Data & Society*, 4(2):205395171774353, December 2017.

[59] Nadezhda Purtova. From Knowing by Name to Personalisation: Meaning of Identification Under the GDPR. *SSRN Electronic Journal*, 2021.

[60] Mark Elliot, Kieron O'Hara, Charles Raab, Christine M. O'Keefe, Elaine Mackey, Chris Dibben, Heather Gowans, Kingsley Purdam, and Karen McCullagh. Functional anonymisation: Personal data and the data environment. *Computer Law & Security Review*, 34(2):204–221, April 2018.

[61] Nadezhda Purtova. The law of everything. Broad concept of personal data and future of EU data protection law. *Law, Innovation and Technology*, 10(1):40–81, January 2018.