# 第一届 数字文明大会学生论文竞赛

# 1st DCC Student Essay Competition

## 论文主题

“在你看来，在未来的数字文明中，谁应该拥有数据？是公司还是个人？或者应当禁止数据共享？请选择你认为最优的解决方案并解释原因。”

## Theme

“In your opinion, who should own data in the future digital civilization? Should companies own data, should individuals own their data or should data sharing be prohibited? Please explain which scenario is better and why.”

DIGITAL CIVILIZATION CONFERENCE 数字文明大会

**Hannah Rose Kirk**
Peking University

"In your opinion who should own data in the future of digital civilization? Should companies own data, should individuals own data or should data sharing be prohibited? Please explain which scenario is better and why."

# KNOWLEDGE AS POWER: A NEW APPROACH TO DATA

## Abstract

This paper argues for a 'right to know' and a 'need to explain' approach to data rights, dismissing the 'right to own' and 'need to sell' approach as inadequate to deal with the complexities of the digital data ecosystem. Purely private or public governance of data cannot contend with data's lack of tangible boundaries, attributable ownership and clarity of use. Thus, data should not be commoditised, but instead knowledge asymmetries in the use of data should be reduced. Through presenting the problems with private and public ownership respectively, this paper finds an explainability approach to data rights offers the greatest flexibility in preserving privacy both across cultures and scenarios. Finally, the discussion of data ownership alone is no longer sufficient to deal with increasingly sophisticated technologies in the form of artificial intelligence or machine learning algorithms which amass and apply data in complicated ways. A knowledge-based approach contends best with protecting data rights also under these scenarios by explaining the data inputs to, and the model output of, algorithmic decisions.

## Who Should Own Data?

"Privacy is the claim of individuals, groups or institutions to determine for themselves when, how, and to what extent information about them is communicated to others."

(Westin 1967,p.7)

Since Westin's first attempt to define individual privacy, the world we live in today as individuals is exceptionally different. The extent of digital traces produced every day across the world has introduced an unprecedented degree of complexity into how we govern the privacy of our online identity. The 'Information Revolution', through inimitable changes to governance, economy and society, has blurred the distinction between private and public ownership [1]. Data's lack of tangible boundaries, attributable ownership and clarity of use has fundamentally changed the tenability of consent on personal privacy. Assigning property rights to data is a fruitless approach; legislating pure private or public ownership benefits neither individual, corporation nor government. Instead, privacy legislation should empower individuals through expanding transparency and explainability of data use, granting the individual not a 'right to own' but a 'right to know' and to understand. A modus operandi focused on knowledge as power, not ownership, better protects data rights today but also offers a more seamless transition by applying the same 'explainability approach' to understanding the 'black-box' algorithms of tomorrow. By eliminating knowledge asymmetries, individuals need not monetise data as an owned, saleable commodity but instead can make autonomous, informed and personalised decisions to govern their own digital future. Each technological iteration has expended the ability to collect more and different forms of personal data, altering our means to protect privacy with respect to changing societal conditions [2]. The camera is cited as an example technology which fundamentally changed the ability to give consent on relinquishing privacy in public places [3]. Conversely, blockchain comprises a new technology offering new ways to protect personal privacy in online spaces. Furthermore, algorithmic decision-making incorporated into healthcare, judicial systems and even social credit scores in China has allowed for the integration of personal data on an unseen scale, tightening the link between our online and offline selves. The future promises to bring increasingly pervasive digital technologies, and whether or not these help or hinder informational privacy, technological advancements change our innate and ontological understanding of what privacy means [4]. Privacy is thus not an invariant or unanimous concept instead varying with time, behaviours, or cultures [5]. Such complexities make privacy as a concept "excessively difficult to define [6]". To protect the right to privacy in the future of digital civilisations, we need a system which respects the complex, changing nature of privacy. Specifically, I propose we need a system which is sufficiently flexible, allowing for different boundaries of privacy across individuals, cultures and technologies, accommodating different types of data inputs and outputs. I argue issuing data property rights, irrespective of owner, cannot simultaneously solve these aims.

A privacy paradox exists, where individuals express concern over protecting digital privacy but do little to keep it safe [7]. In an article for the Economist, musician Will.i.am lamented the exploitative power of the 'data monarchs', considering instead "data itself should be treated like property and people should be fairly compensated for it" [8]. There's a problem here; Property contracts were designed to deal with tangible, contestable assets like houses, cars or farmland [9]. Data is not a commodity, it's information and so gives traction to the privacy paradox, where we cannot protect our data even if we wish to because it's intangible, invisible and increasingly diverse. Assigning a single owner of data is infeasible, we don't own our data, nor do companies or governments. The interconnectedness of data collection, specifically from the digitalisation of social interactions, has muddied ownership yet further. Consider two individuals conversing over Facebook or WeChat, if each individual owns the data, which takes priority in choosing whether to share the conversation's content? A shift from personal autonomy to co-dependence was pertinently demonstrated by Cambridge Analytic, who were able to collect information on up to 50 million users with permission from just 270,000 [10]. The security of our data no longer depends on our own actions but on those beyond our u control. The privacy paradox exposes the difficulties to define data as a possession, refuting ownership rights as a tenable solution.

Even if individuals could claim data ownership, the uncertain value of data and large volume of data existing online only works to reduce the payment we would receive [11]. Admittedly, the exploitation of data has generated enormous fortunes for a handful of companies which together exceed $3 trillion in market value. Yet, even if we follow the plea of California Governor Gavin Newsome [12], in creating a data dividend to share out this generated wealth, the amount received per individual pails in comparison to the aggregate fortunes. Each of Facebook's 2 billion users would receive only $9 a year if profits were proportionally redistributed [13]. The pitiful worth of personal data was rudely brought to attention when it was revealed Cambridge Analytica paid only $0.75 per US voter profile for in-depth information spanning surname, gender, location, personality traits and political preference scores [14]. Irrespective of monetary value, the question remains whether the right to own data and the concomitant right to sell information about ourselves is a desirable treatment of personal privacy. Basing privacy protection on property rights commodifies data and introduces a financial dependency to earn from it. This 'data prostitution', where we lack autonomy to decide what

we do with our data but instead seek to monetise it, invokes serious problems. Data inequality threatens societal fairness in digital civilisations where less advantaged groups pay their way with information, and only the rich can afford the right to privacy. We can, and already do, pay with our data without it being monetised explicitly, by exchanging information for a service. When Google's unfamiliar business model first emerged, the offering of free services appeared benevolent. It was a case of too good to be true, European competition commissioner Margrethe Vestager, recently stated, "this idea of services for free is a fiction… people pay quite a lot with their data for the services they get" [15]. By recording our preferences and that of others, Netflix can recommend new favourite films; by matching our profile to the demands of hiring firms, LinkedIn can help us find a dream job; by tailoring adverts to our behaviours, e-commerce platforms reduce knowledge asymmetries and search frictions. Few would readily give up the private benefits we gain from these services in lieu of data protection. In fact, many of us readily click through Terms and Conditions to use free services without reading what permissions we are giving away. Data ownership is too simplistic, being paid a few pennies for our data would unlikely change such behaviours and could introduce considerable data inequalities both domestically and globally.

If privatising data ownership is an unsatisfying solution, the hailed alternative is public data sharing. Data is intangible and can be used by many people simultaneously so leans towards the 'information as a public good' argument [16]. The non-rivalry, non-excludable nature of data creates a 'system resource' involving people, platforms and profiles [17]. Data sharing appeals by amassing public value, and monetising data would disrupt such important use [18]. International development and humanitarian organisations are increasingly calling for digital data to be treated as a public good because of its value in supplementing scarce national statistics and informing interventions, including in emergencies [19]. Large-scale datasets of call detail records supplied by telecom giants are being used to target the spread of diseases, such as dengue [20] and cholera [21], and to map poverty [22] and mobility [23]. Despite the power of data sharing for societal benefit, issues remain. Public ownership of data or the right for corporations to sell on data conflict with an individual's right to privacy, even if it is for communal social benefit. Sharing call data presents a high risk to sensitive information, and even de-identified data could allow people to be tracked not just for diseases, but also for political interest or dissident behaviours [24]. We might accept some level of data tracking, trading personal privacy for a tightening of social security, but mass surveillance remains unacceptable. As such, public ownership has equally as many failings as private ownership in dealing with the necessitated flexibility of digital governance to many preferences and scenarios. Instead, we need a system which permits sharing personal information that is useful for our social, economic, and governmental systems while still protecting vital personal autonomy.

Granting consensual data rights delivers a tenable resolution to such complex discussions of data ownership. Geoffrey Canright of Telenor ascribes to the centrality of knowledge in the process of data sharing, "'informed consent is understanding what you are saying yes to, and knowing what data are being used and why. You can take your consent back at any time [25]". Treating knowledge as power has gained traction in international privacy legislation. Germany has a legal concept known as "informationelle Selbstbestimmung", under which an individual has the right to decide for themselves what information can be used by whom and for what purpose. The European Union's GDPR acts as a benchmark for data rights, giving individuals the personalised authority to decide on data usage within their privacy boundaries. The regulation maps data flows within organizations, enforcing the creation of back-end systems which enable individuals to access, correct and delete information. One year on from the GDPR, China introduced its own, albeit more permissive, data regulations of this kind. On May 28th of this year, the Cyberspace Administration of China demanded customised content using recommendation algorithms driven by personal information, including news feeds and advertising, should be explicitly labelled. Instead of focusing on ownership, we should divest our efforts to closing knowledge asymmetries between what corporations and government know and what individuals don't know. To re-empower individuals and to share knowledge between all actors in the digital ecosystem, privacy law should directly give individuals, as the generators of data, the right to know [26] and enforce corporations and governments, as the users of that data, a responsibility to explain. This principle of transparency, a central tenet of the GDPR, best serves the interest of individual data privacy without stemming the social benefits from data sharing.

With increasingly sophisticated technologies penetrating our lives with unprecedented precision, we must go beyond considering the applicability of digital governance strategies to today's data infrastructure. Instead, any regulations or best practices require adaptability and agility to new relationships with technologies facilitated by machine-learning and artificially intelligent systems. Many cite China's advantage in AI as stemming from its data resources, the oil of today's industrial revolution [27]. Concerningly, when data rights are presented as an

obstacle to technological innovation, a race to the bottom of data exploitation ensues, where individuals inevitably lose. Consequently, we need a system of data rights which equally introduces transparency into future technologies. Currently, 'black-box' systems such as neural nets, support vector machines and matrix factorisation provide little interpretability of what data inputs are used and how outputs are reached [28]. White-box methods, such as decision-trees, are more interpretable but at the cost of lower accuracy. A lack of explainability has serious consequences when considering the mass applications of these algorithms to healthcare, to judicial systems or even to the deployment of autonomous weapons. We may trust algorithmic decisions relying on logic obscured by 'hidden layers' for recommender systems of movies or commercial products, but the stakes are higher when we consider systems such as China's proposed social credit system which have tangible impact on individuals' offline lives. We can transfer the current argument of 'the right to know' to these technologies in three steps. Firstly, to understand the technical model at hand, granting individuals an understanding of its opacity. Second, reveal the data inputs the model is trained on. Finally, disclose how the model will be used in practice. Knowledge-dependency graphs [29] or LIME [30] are steps in the right direction to create interpretable models upon which Informed consent can be reintroduced. Articles 13-15 in the GDPR mandate a 'right to explanation', a remit which offers consistent protection of data privacy withstanding complicated technological advancements. Data ownership once again falls short, because under more intelligent machine learning algorithms, even if we understand what data inputs we forego or sell, we do not necessarily understand how our data is being processed.

At current, for most of us, lapses in privacy protection cause manageable annoyances in daily life, like following a search of lightbulbs on google, we cannot escape lightbulb ads littering our social media, or after buying one product on Taobao or Amazon we are repetitively shown similar product despite only wanting to buy one. Yet data misuse and exploitation can have serious consequences at the detriment to individuals and also to society at large. The manipulation of public opinion, the dissemination of fake news and the potential for discrimination through mass surveillance are not mere annoyances. Treating our data as property has appeal, especially when such scandals damage trust individuals place in corporations or governments. Nevertheless, data does not lend itself to ownership, both due to its low monetary value to the individual and high social value to others. Data should not be commodified by the 'right to own', but its uses should be explained in advocacy for a 'right to know'. By presenting the failings of private and public ownership, it becomes clear the greatest suitability to the most scenarios is offered by granting data rights to ensure data is stored, processed and protected in a trusted environment but its uses remain open to scrutiny. Accordingly, we should build technologies with these tenets in mind; distributed ledgers, trust encrypted environments and explainable 'white-box' algorithms are technical solutions which could be used to aid the enforcement of promises from policies like GDPR. Applying the language of blockchain, individuals, as decentralised citizens, can be empowered to know, correct, and delete personal information about themselves, and companies are legally required to offer immutable, incorruptible explanations. It follows that any discussion of digital governance must consider how privacy can be protected in a world where data is gathered and shared with the increasing speed and ingenuity introduced by artificially intelligent systems. Relying on knowledge of data inputs and explainability of algorithmic outputs consistently safeguards privacy in the future in a more seamless transition. By assimilating sufficient information for consensual choice, personal privacy is protected, removing the frictions introduced by ownership and avoiding strangling societal data sharing in the public interest. In answering this complex question, knowledge is power, and applying this logic to data rights will ensure a mutual beneficially digital future for us all.

DIGITAL CIVILIZATION CONFERENCE 数字文明大会

## ENDNOTES

[1] Matthews, 2000

[2] Westin, 2003

[3] Holvast, 2007

[4] Floridi, 2005

[5] Margulis, 1977

[6] MacCormick, 1974 (p.75)

[7] Barth and Jong, 2017

[8] Will.i.am, The Economist, January 21 2019

[9] Peck and Shu, 2017

[10] B Stewart, Vox. March 20 2018

[11] Gierl and Huettl, 2010

[12] Thompson, AP News. February 14 2019

[13] The Economist, July 7 2018

[14] Hill, The Register, March 30 2018

[15] Walt, TIME. November 8 2018

[16] First proposed by economists Stiglitz and Varian.

[17] Putova, 2015

[18] For example, because Census data is freely available as part of the information commons, it forms the backbone of important policy research.

[19] Taylor, 2016

[20] Wesolowski et al., 2015

[21] Bengtsson et al., 2015

[22] Gutierrez et al., 2013

[23] Taylor, 2016

[24] Taylor, 2016; Berlingerio et al., 2013

[25] Geoffrey Canright. Head of Data Analytics Group, Telenor, Sep 2015 (from Taylor, 2016)

[26] An Ipsos global survey exposed the scale of these asymmetries, where two-thirds of respondents said they knew little or nothing about how much data companies held about them or what companies did with their data, CIGI-Ipsos Global Survey on Internet Security and Trust 2018

[27] The Economist, May 6 2017

[28] Abdollahi, 2017

[29] Alshammari et al., 2019

[30] Local Interpretable Model-Agnostic Explanations, Ribeiro et al., 2016

## REFERENCES


Abdollahi, Behnoush. 2017. “Accurate and Justifiable: New Algorithms for Explainable Recommendations.” University of Louisville.

Alshammari, Mohammed, Olfa Nasraoui, and Scott Sanders. 2019. “Mining Semantic Knowledge Graphs to Add Explainability to Black Box Recommender Systems.” IEEE Access 7: 110563-79.

Barth, Susanne, and Menno D.T. de Jong. 2017. “The Privacy Paradox-Investigating Discrepancies between Expressed Privacy Concerns and Actual Online Behavior-A Systematic Literature Review.” Telematics and Informatics 34 (7): 1038 - 58.

Berlingerio, Michele, Francesco Calabrese, Giusy Di Lorenzo, Rahul Nair, Fabio Pinelli,and Marco Luca and Marco Luca Sbodio. 2013. “AllAboard: A System for Exploring Urban Mobility and Optimizing Public Transport Using Cellphone Data.” In, 663-66.

CIGI-Ipsos Global. 2018. “Survey on Internet Security and Trust.”

Floridi, Luciano. 2005. “The Ontological Interpretation of Informational Privacy.” Ethics and Information Technology 7 (4): 185-200.

Gierl, Heribert, and Verena Huettl. 2010. “Are Scarce Products Always More Attractive? The Interaction of Different Types of Scarcity Signals with Products’ Suitability for Conspicuous Consumption.” International Journal of Research in Marketing 27 (3): 225-35.

Gutierrez, Thoralf, Gautier Krings, and Vincent D. Blondel. 2013. “Evaluating Socio-Economic State of a Country Analyzing Airtime Credit and Mobile Phone Datasets,” September.

Hill, Rebecca. 2018. “$0.75-about How Much Cambridge Analytica Paid per Voter in Bid to Micro-Target Their Minds, Internal Docs Reveal.” The Register, March 30, 2018.

Holvast, Jan. 2007. “History of Privacy.” In The History of Information Security, 737-69. Elsevier.

Kamleitner, Bernadette, and Vincent-Wayne Mitchell. 2018. “Can Consumers Experience Ownership for Their Personal Data? From Issues of Scope and Invisibility to Agents Handling Our Digital Blueprints.” In Psychological Ownership and Consumer Beha vior, 91-118. Cham: Springer International Publishing.

Mathews, Jessica T. 2000. “The Information Revolution.” Foreign Policy, no. 119: 63.

Purtova, Nadezhda. 2015. “The Illusion of Personal Data as No One’s Property.” Law, Innovation and Technology 7 (1): 83-111.

Ribeiro, Marco Tulio, Sameer Singh, and Carlos Guestrin. 2016. “‘Why Should I Trust You?’: Explaining the Predictions of Any Classifier,” February.

Stewart, Emily. 2018. “Facebook’s Cambridge Analytica Crisis Keeps Growing.” Vox, March 20, 2018.

Taylor, Linnet. 2016. “No Place to Hide? The Ethics and Analytics of Tracking Mobility Using Mobile Phone Data.” Environment and Planning D: Society and Space 34 (2): 319-36.

The Economist. 2017. “The World’s Most Valuable Resource Is No Longer Oil, but Data,” May 6, 2017.

The Economist. 2018. “Data Workers of the World, Unite,” July 7, 2018.

Thompson, Don. 2019. “California Governor Wants Users to Profit from Online Data.” AP News, February 14, 2019.

Walt, Vivienne. 2018. “She Took on Silicon Valley Giants. Now Margrethe Vestager Is Preparing for Her Final Act.” TIME, November 8, 2018.

Wesolowski, Amy, Taimur Qureshi, Maciej F. Boni, Pål Roe Sundsøy, Michael A. Johansson, Syed Basit Rasheed, Kenth Engø-Monsen, and Caroline O. Buckee. 2015. “Impact of Human Mobility on the Emergence of Dengue Epidemics in Pakistan.” Proceedings of the National Academy of Sciences 112 (38): 11887–92.

Westin, Alan F. 2003. “Social and Political Dimensions of Privacy.” Journal of Social Issues 59 (2): 431-53.

Will.i.am. 2019. “We Need to Own Our Data as a Human Right—and Be Compensated for It.” The Economist, January 21, 2019.


## Thank you.