Alternative Personal data governance models

Bastian Greshake Tzovaras^{1,2}, Mad Price Ball²

¹Center for Research & Interdisciplinarity (CRI), Université de Paris, Paris, France

²Open Humans Foundation, USA

This text is a draft of a chapter for the forthcoming book Automating the Crowd, which is planned to be published in 2020. This manuscript is likely to be changed and updated as the book is edited in context of the other chapters.

Automating the Crowd will unpack the relationship between the work done using AI and the humans who make it work. It will weave together the untold stories of the workers and volunteers whose effort powers AI systems with discussions of the impact of those systems on society. This edited collection will address the technical, social and ethical implications of the use of 'crowdsourced' efforts and the importance of increasing diversity of contributors in human-machine interactions.

Automating the Crowd was developed at an event at the Alan Turing Institute in London in January 2019 (<https://www.turing.ac.uk/events/automating-crowd>) and a follow up discussion in Cambridge in July 2019 (<http://www.crassh.cam.ac.uk/events/28618>). It brought together participants from neuroscience, digital humanities, sociology, health and other disciplines to shape the public debate around the relationship between automation and human work.

If you would like to be notified when the book is ready to be published, you can subscribe to its email newsletter at <https://tinyletter.com/automatingthecrowd>.

Table of Contents

[Framing the problem: Incidents and existing gaps in oversight](#)

[The YFCC100M image data set: altruistic sharing & unanticipated use](#)

[Questionable uses of AI/ML](#)

[Regulation and oversight of AI/ML uses](#)

[Governance models for personal data](#)

[Consent-based approaches and limitations](#)

[Biobanks as historic precedent](#)

[Personal Data Exchanges](#)

[Privately operated exchanges using broad consent](#)

[Privately operated exchanges using individual consent](#)

[Collectively operated exchanges using broad consent](#)

[Collectively operated exchanges using individual consent](#)

[The value of personal data](#)

[Trust in Personal Data Exchanges](#)

[Governance for uses of data instead of access of data](#)

The not-so-secret ingredient that underlies all successful Artificial Intelligence / Machine Learning (AI/ML) methods is training data. There would be no facial recognition, no targeted advertisements and no self-driving cars if it was not for large enough data sets with which those algorithms have been trained to perform their tasks. Given how central these data sets are, important ethics questions arise: How is data collection performed? And how do we govern its use?

Framing the problem: Incidents and existing gaps in oversight

While there have been concerns about the potential consequences of AI/ML technologies for decades, recent incidents help provide concrete examples of why governance is desired. These incidents often involve unanticipated uses of large datasets, and may be producing uses that are harmful and/or repugnant to individuals from whom data was collected.

The YFCC100M image data set: altruistic sharing & unanticipated use

One source of ongoing lessons is the release and uses of the *Yahoo-Flickr Creative Commons 100 Million* (YFCC100M) image data set that was created by Yahoo in 2014. As the name suggests, it contains nearly 100 million images that Flickr users released under a *Creative Commons* license, around 68 million of which are accompanied by have user-generated tags, such as *baby, cat, dog, park, nature*. Making this data set broadly available was intended to benefit the AI/ML community by sharing an invaluable resource for developing algorithms that identify the content of images. More recently, IBM released a derivative data set in 2019, called "*Diversity in Faces*" [2], annotating 1 million human facial images with objective measures (e.g. craniofacial features) as well as subjective (e.g. age and gender classification via human annotators).

The data sets are noteworthy due to the altruistic intents in sharing done by both the original sources (photographers sharing with *Creative Commons* licenses) as well as the data set curators (Yahoo and IBM). The re-uses by the latter were almost entirely unanticipated by the photographers that shared their work – and so it is not surprising that many had concerns about the release. Despite the broad licensing on the photos, reporting by *NBC News* framed this as a consent issue: "*Facial recognition's 'dirty little secret': Millions of online photos scraped without consent*" [3], and described concerns about further unanticipated uses – e.g. potential use of resulting AI/ML to target minority groups.

Questionable uses of AI/ML

The potential for AI/ML to harm particular demographics demonstrates how concerns about AI/ML extend beyond risks for individuals within the data itself – to broad questions of potential social harm. Similar concerns were made more tangible when a 2017 study by Stanford researchers trained an artificial intelligence to distinguish between the faces of gay and heterosexual men and women. The findings sparked outcries by LGBTQ civil rights group [4], as well as concerns about consent and the claimed "public" nature of the data – images and data scraped from over 75,000 users from a dating website.

We know about these AI/ML uses because they are published – but the lion's share of AI/ML use is in private contexts, where "big data" holders (e.g. Google and Facebook) use the personal data we provide to train their machine learning-based products [5]. Reading through all of our emails and search queries to give personalized, predictive auto-replies as well as more targeted advertisements, tracking our interactions with friends to market our interests. We typically "consent" to our data being re-used via "Terms of Service", but these terms are rarely carefully read – indeed, it's been estimated that actually doing so for each site visited over the course of a year would take 30 to 50 minutes per day [6]. Even when these terms are read, exact uses are unlikely to be disclosed, and our agency is limited by an absence of alternative services – one study discovered that "quitting Facebook for a year" requires an incentive of over \$1000 USD [7].

Regulation and oversight of AI/ML uses

Academic research and private uses of AI/ML highlight overlapping issues regarding gaps in regulation and governance of AI/ML applications to data. Broadly speaking, there are two junctures at which decisions can influence this process: (1) the data collection phase, where there is an opportunity to establish consent and governance for the data set, and (2) the data use stage, where review can determine whether a specific use of AI/ML should occur. The second juncture – when the data is used – has historically had little oversight beyond any review process or agreements (if any exists) required to receive access to data.

In private contexts, companies engaged in AI/ML, "big data", and web technologies have access to comprehensive data which are collected by their platforms (and are thus minimally subject to data use agreements). Recently, companies are increasingly under pressure to address ethical concerns in their uses of these data, resulting in some companies creating "ethics" positions (e.g. "Chief Ethics Officer") responsible for implementing oversight. Even for companies doing this, however, there is cause for skepticism: the results may be performative (e.g. simply satisfying procedure checklist), and even those procedures may be poorly designed – prone to "copycat" and imitating popular ideas, rather than specifically addressing the risk landscape of their own work [8].

In academic contexts, a review of AI/ML applications has been minimal or absent when data sets are "public" or "anonymous", as this removes them from ethics review requirements. However, this approach is flawed, based on a historic assumption of "low risk" that no longer hold true, as captured by the Council for Big Data, Ethics & Society:

At a prior point in history, it was reasonable to assume that *publicness* of an existing dataset was an adequate proxy for informational risk. Because disclosure of the data itself was the relevant harm, it was safe to assume that if data was already publicly available then by definition any harm was already done. However, the power and peril of big data research methods is that large datasets can theoretically be correlated with any other large datasets to algorithmically discover patterns that can re-identify and/or impact individuals in ways that were previously impossible.

– Metcalf, Jacob. 2019. "Letter on Proposed Changes to the Common Rule."
Council for Big Data, Ethics, and Society. Accessed September 18, 2019. [9]

Furthermore, the academic context of ethics oversight relates to human subjects research and consequently concerns itself with risks posed to individuals within a data set – and not regarding the broader societal impacts of data use.

Taken as a whole, shortcomings in oversight on AI/ML uses of data leads many to consider what influence can occur at the first stage – data collection – via new approaches for consent and governance of resulting data sets.

Governance models for personal data

Separate to oversight of AI/ML applications themselves, there are a variety of approaches for in governance on personal data, for making decisions regarding allowed uses and sharing. Historically, there has been heavy reliance on "up front" decisions made by individuals to guide later data use, via "terms of use" or "informed consent". This format is increasingly recognized to have challenges, and a variety of additional approaches for governing data access have arisen in efforts to address these issues.

Consent-based approaches and limitations

Historically, the different ways that data are allowed to be used have been established in one-off decisions: for example by the "terms of service" that users of an app or website agree to, or the "informed consent" of participants in research. This last scenario – "informed consent" in research – has a history that predates the internet and digital research, and draws heavily on experiences in biomedical research and "biobanks" that store biological materials. These experiences are valuable context: they represent an area where researchers and ethicists have cooperated heavily, and thus represent earnest efforts which reveal fundamental limitations in a purely consent-based approach. Lessons from biomedical efforts for consent are valuable insights on use of "terms of use" and "privacy policy" in the broader landscape of personal data governance.

Improving comprehension is difficult, especially when consent is performed as a comprehensive, one-off decision. Individuals are expected to review explanations for diverse potential future scenarios, and then make an informed choice. Research into improving informed consent forms and processes provide disheartening results: the amount of information is challenging, people skim and skip text forms, and efforts at simplification and engagement are challenging. Furthermore, no amount of simplification can remove the cognitive process of considering risks and benefits. Many studies attempting to simplify and clarify consent forms have failed to find significant improvements in comprehension [10, 11]. Outside research, we can be certain that the situation is even worse: one study estimated that, in theory, reading all privacy policies would take an individual 244 hours (30 workdays!) per year – something few can afford to do [6].

"Granular consent" is an approach to make consent less monolithic. A monolithic "agree to everything" approach disempowers individuals, especially when there are no practical alternatives ("take it or leave it"). "Granular" approaches for consent enable individuals to select which (and not all) items they agree to, and are more readily implemented when consent processes are themselves digital. Reflecting the importance of this approach the EU GDPR calls for separate consent for distinct processing activities [12] – in practice, however, activities are not obviously distinct, and consent may remain largely monolithic.

"Broad consent" is a contrasting approach, motivated by a desire to enable beneficial and wanted uses of data – uses that are unanticipated, and which may be prevented by a narrowly scoped agreement. Previously, these required costly processes that made re-use infeasible: re-consent and/or new IRB approval. In some frameworks, data holders might be able to seek new consents for new use cases, but this relies on participants being available and willing to engage in additional review and authorization. The shift to "broad consent" was motivated by the opportunities of "big data" technologies to perform discovery on aggregate genomic and health data [13], and reflects concerns about the potential social costs of specific authorization – proposing to instead trust in the decision-making of resource stewardship.

The use of "broad consent" brings its own problems, and may ultimately be equivalent to the gaps in oversight that have already existed for other data sets: unanticipated uses have little to no oversight, and may be problematic or concerning. Indeed, a controversial incident in genomics seems to paint a very similar picture. In August 2019, researchers used data from the Direct-To-Consumer genetic testing company *23andMe* and the *UK Biobank* to investigate *"the genetic architecture of same-sex sexual behavior"*. In the ensuing post-publication discussion, especially the broad consent of the *UK Biobank* came under scrutiny [14], as participants of it consent to have their data used *"to improve the prevention, diagnosis and treatment of illness and the promotion of health throughout society"*. Whether investigating the existence of the *gay gene* is covered by this consent is debatable, especially in the light of the potential harms to the LGBTQ community that could result from such research.

Regardless of the approach, when consent-based governance relies on a singular agreement event to "set the rules", it relies heavily on our collective abilities to anticipate future use – by consenting individuals, and

by those who make the requests. This can fail in both directions: by not allowing for unanticipated "good use", and by failing to prevent "bad uses". Either way, individuals are likely unaware of what they have (and haven't) agreed to.

Biobanks as historic precedent

Biobanks share many of the same challenges as personal data as they manage consent, sharing, allowed use of resources from individuals. As such, biobanks serve as a forerunner in exploring alternative governance models, particularly in the form of patient-led biobanks..

While most biobanks are collected and run by research and medical institutions, patient-led biobanks have demonstrated the ability for some self-governance in managing a research resource derived from a community. One early example of this was a patient-registry and biobank that was started in the 1980s by families affected with Canavan-disease – a degenerative disorder that causes progressive damage to nerve cells in the brain. By providing researchers with access to data and biological samples, they could steer the direction of research and incentivize research. The patient group could even defend their rights in court, when a researcher who was given access to the samples tried to patent a test for Canavan disease.

A similar model was adopted by the *Genetic Alliance Registry and BioBank* (GARB) which was founded in 2003. Following the model of the *Pseudoxanthoma Elasticum International Registry and BioBank*, six different disease advocacy organizations came together in an effort to collect data and samples from their respective members. By building a shared infrastructure for their registry & biobank, the members of GARB could help to facilitate and accelerate translational research.

Biobanks managed by patient advocacy organizations serve as early examples for alternative governance models – but questions then arise regarding the governance of those organizations. While patient-advocacy organizations are hopefully more aware of patients' wishes and needs and thus better aligned with them, it remains unclear of how representative they are for the larger community of contributors to a given biobank. Furthermore, it remains unclear of what influence an individual participant in a biobank can have over shaping the governance of the patient-advocacy body that runs the biobank itself.

Personal Data Exchanges

Just as the traditional model of Biobanks has evolved to enable alternative governance models around biological samples, there is an increasing interest in designing and building alternative governance models for personal data. And there is a growing diversity in approaches in how to such governance models around personal data can look like. Unlike Biobanks – which have approached a collective solution to governance with a patient board making the decisions – many of the alternative governance solutions for personal data attempt to put the data governance into the hands of the individuals. In some cases, platforms seek to perform data monetization and return value to individuals. Broadly, we call these approaches for individuals' managing permissions for aggregated personal data Personal Data Exchanges (PDEs) [15].

As a technology platform the goal of PDEs is to support active decisions made by individuals to share their data. PDEs commonly support for individual-centered aspects like agency in control or sharing decisions, privacy management, and/or distribution of value – which often leads to self-characterization as enabling data "ownership". By doing this, these exchanges try to change who is in control, shifting ownership from companies at least partially to the individuals who generate the data. Over the years a growing number of such PDEs have launched, often being differentiated by slightly different business models behind them and a different focus on the types of data they want to aggregate.

These business models and approaches can be differentiated on two different dimensions. Firstly, governance of the operating entity and infrastructure used for the PDE: is it a private entity or collectively owned by participants? And secondly, governance of how data sharing decisions are made: as individual consent for different data use cases or through broad consent? We will examine these four different approaches by looking at existing Personal Data Exchanges that try to implement such models to examine how such implementations for alternative data governance can work in practice.

Governance of....		Data sharing decisions	
		Broad consent	Individual consent
Operating entity	Private	<i>Datacoup</i>	<i>Digi.me</i>
	Collective	<i>LunaDNA</i>	<i>MIDATA</i>

Privately operated exchanges using broad consent

Frequently such PDEs are run as private companies, which own and run the technology to facilitate the data exchange itself. An example such PDE is *Datacoup*, which aims to get individuals to aggregate their personal data across different web platforms in their own system. In case of *Datacoup*, individuals are encouraged to import their data into the *Datacoup* system where it will then be sold by *Datacoup* to buyers in the form of de-identified aggregates or derivative analytic information. *Datacoup* also provides an individual consent model for selling personal data directly to third parties [16]. In return for participating in the broader analysis platform, *Datacoup* returns value to individuals based on the revenue generated for that category of data [17].

The financial success of both broad and individual consent approaches for *Datacoup* remains unclear: there has been little news in recent years following its last round of funding in 2015 [18]. The private governance of the organization also implies that, if it is successful, it could take an arbitrary amount of profit with minimal return to individuals. Furthermore, this lack of collective governance may raise concerns regarding the uses of personal data, as the broad consent means that the company providing the PDE has the right to engage in diverse data uses at the behest of 3rd party interests.

Privately operated exchanges using individual consent

Other privately-owned data storages and exchanges, such as *Digi.me*, follow a slightly different model. Similarly to *Datacoup*, they encourage individuals to import their personal data ranging from health care to financial data. Instead of giving a broad consent to the data use down the line, users of *Digi.me* can consent to individual use of their data through an app store-like model. Apps built around *Digi.me* can then access the user's private data sets. Currently, most of these apps are targeting the *digi.me* end users, to provide services such as financial dashboards, analyses of their social media profiles or health recommendations. Unlike *Datacoup*, there seems to be no model in place for the individual members to earn money through sharing their data at this point. Instead, at this stage the business model of *digi.me* relies on the app developers paying *digi.me* for using the data stored in their data store.

Setups like these help to alleviate some of the problems around broad consent by giving the individual more control over how and when to share their personal data with other parties. The larger issue of users having control over the data exchange/store themselves remains though. As the platform itself is held privately, they can not control the direction the platform will take. As a result potential monetization strategies might change down the road, with users having no control over this, except for deleting their data and revoking consent.

Collectively operated exchanges using broad consent

Different approaches have been implemented to get around the limitations of purely privately owned Personal Data Exchanges. One effort that describes itself in terms of "community ownership" is *LunaDNA*, which specializes in being a PDE for genetic data and health records with the goal to sell access to these data to researchers. Individual users of the system consent for their data being used by LunaDNA and 3rd party researchers as part of the aggregated data across all users. Additionally, users can opt-in into providing more data for specific research questions. In return for allowing the company to share their data in return for income, users receive dividends based on their data contributions.

LunaDNA describes itself as engaging in community ownership: "*Community ownership means people who join and share health data can claim shares of ownership in LunaDNA*". LunaDNA, LLC is managed by the Luna Public Benefit Corporation, a business model which remains for-profit but mandates some public good within a company charter. Despite these claims to community ownership, it is unclear to what extent community should be considered "owners" with any governance role: while the managing "public benefit corporation" charter has some mandate for public good, members are almost entirely excluded from any governance of the company itself, as is explicitly recognized in its SEC filings.

Collectively operated exchanges using individual consent

Another approach to shared ownership has been implemented by the MIDATA project, which aims to develop national cooperatives which are to host the Personal Data Exchange. The project is being developed in Switzerland, where the first cooperation has already launched. While users are not required to become members of the cooperative, they can do so to gain voting rights at the cooperative's General Assembly. Using their voting rights, members can elect the administration of the cooperative, vote on amendments to the Articles of Association and MIDATA guidelines, as well as decide on how to use potential profit. For the Swiss cooperative, the application to become a member is tied to a one-time fee of 40 CHF.

This overall model is designed to give cooperative members a high degree of influence over the governance of the PDE. Due to legal restrictions, cooperative membership is limited to residents of the country in which the cooperative operates (i.e. Swiss residents for the Swiss cooperative). Unlike the broad consent approach of LunaDNA, users of the MIDATA Personal Data Exchange can import data to share them with specific research studies and projects, to which they consent on a case-by-case basis. In the current iteration of the MIDATA system, there are no plans to create a profit from the data collected by the users. Instead, the individual cooperatives are non-profit organizations which do not pay dividends to the cooperative members or financial compensation to the users that have donated the data.

Another example of PDE in this format is one led by the authors of this chapter, *Open Humans*, which originated as a similar effort to enable individuals to share diverse data for potential research uses. Also operated as a nonprofit, the platform supports collective input into platform governance via community-elected seats on the Board of Directors, as well as via consensus-based review by the community to determine whether projects may be made broadly-available to the membership. While these models for collective governance may ensure a stronger voice for the platform's users, they also face

challenges of a nonprofit context: developing and maintaining these platforms is nontrivial, and more challenging when it relies on volunteer labor, contributions, and grants.

The value of personal data

While not all data exchanges try to offer a direct monetization of the user's personal data, virtually all of the exchanges share the idea that personal data does have value to third parties. It is hard to understand how much value an individual data set should have though, with estimates fluctuating wildly from a few dollars up to a \$100 for the data of a Facebook profile [19]. LunaDNA, for example, estimates that their users might be able to get shares worth at around \$21 [20]. What is clear, is that it is challenging to obtain financial value from individual data sets: LunaDNA's SEC filings, for example, estimate the value of whole genome data to be a mere \$21 – data which costs from \$600 to \$1000 for consumers to obtain.

Achieving value and increasing it further seems to require achieving some sufficiently large aggregate: 23andme achieved large research deals with pharmaceutical partners only after having collected data on 800,000 customers [21]. As a result, there is an inherent trade-off between maximizing the amount of aggregated data and individual consent models that give people more control over to whom to sell the data. Thus, it may be unsurprising that Personal Data Exchanges which aim to earn profits through selling data – such as LunaDNA and Datacoup – favor a broad consent model to maximize the number of users and thus data sets from which they can resell de-identified aggregates and data analyses to third parties.

On the other hand, PDEs that do not focus on the monetization aspect of the data can favor individual consent models for their users as they have to rely less on the maximization of their data sets. As a consequence, these systems enable users to pick and choose to whom to give access to their data. Based on these trade-offs it becomes clear that there might be no one-size-fits-all model for personal data exchanges when it comes to the consent they utilize. Rather, the right model needs to depend on both the incentives/motivations for why users want to share their data as well as on the overall goal of the exchange.

Trust in Personal Data Exchanges

Regardless of the consent model that is adopted, users of personal data exchanges need to trust that their data is stored, managed, and used appropriately by the body that is running the exchange. On a technical level, the data storage needs to be done securely to ensure that data will not be breached or accidentally leaked. This in itself is already not an easy task, as the website *Have I Been Pwned* – which lists data breaches – demonstrates: Their list includes over 400 breached websites including those for large corporations such as Comcast and Adobe, totalling over 8 billion compromised user accounts. Given that *Personal Data Exchanges* are supposed to aggregate personal data from a variety of sources, the potential damages through leaked data might consequently be even more severe for the users of a compromised exchange.

Besides the technical risks that can hinder trust, there is also the aspect of governance: how are decisions made by the managing entity?. This becomes especially relevant in the context of PDEs which are run by privately owned companies, as participants of the exchange do not have any control over how the resource is managed. Thus, even exchanges which are currently managed in line with the expectations and wishes of the user community might change with whom they are doing business, pivot their business model at some point along the way or might even be acquired another company outright. Far from being hypothetical, the past years have seen examples this: In 2018, Google not only quietly remove it's former motto "*Don't be evil*" but also participated in a US military program on using AI systems to analyze drone footage [22]. And of course there is the case of Cambridge Analytica, which acquired data on tens of millions of Facebook users, leading the public trust in Facebook to drop by 66% [23]. Given these risks, people might

rightfully be skeptical on whether they want to deposit copies of all or most of their personal data in a privately held *Personal Data Exchange*.

Governance for uses of data instead of access of data

As we have seen, personal data exchanges try to address the problems around data access and the governance around access to data through a variety of mechanisms, utilizing either broad or individual consent models, in which individuals make the decision on with whom to share their data. But neither of these approaches, whether broad or narrow consent and privately or cooperatively owned, provides a silver bullet around the problem of how the data can be used once it is shared.

The consequences of the YFCC100M data set and potential uses of that data provide one example: as a technology, uses of facial recognition technology is under increasing scrutiny. Concerns include the potential impact on exacerbating inequity (e.g. due to algorithmic errors made more often for certain demographics), privacy invasion, and a threat to civil rights. At the time writing this chapter, in 2019, there have been calls made by researchers [24] and advocacy groups for moratoriums on the use of facial recognition technologies, and regulations on their use.

There is already existing precedent for regulations that strive to protect people from discrimination on the basis of personal data. In the field of genetic data, the United States the 2008 Genetic Information Nondiscrimination Act prohibits the use of genetic information as the sole basis for health insurance coverage and rates, as well as prohibiting the use of genetic information by employers in hiring and job placement decisions.

In France, a judge has ordered a 5-year ban on analysis of court records data for the purpose of evaluating and predicting judicial practices [25]. The ruling came as a reaction to unanticipated consequences of public court data and concerns about the use of analytics to customize legal approaches, and thereby manipulate judicial decisions. Critics of this recent decision raise concerns that it is overly broad [26]; one potential harm is that it insulates judicial practices from public scrutiny – for example, a recent analysis of US judicial decisions which revealed a troubling level of arbitrariness in asylum rulings [27].

While the scope and merits of such regulations are subject to debate, the need for them may be inevitable: an increasing amount of data is becoming available for diverse uses. These may occur in contexts where the ethical validity of individual consent is rendered meaningless due to lack of comprehension, lack of alternative options, and inability to anticipate all potential uses of data. In many cases, consent may not exist at all: data may be collected in public contexts and by government entities.

The consequence of AI/ML inevitably moves our questions from an individual model to a collective one, forcing us to ask not merely "who is allowed to use our data?" but also: "what are they allowed to do with it?"

References

- [1] *YFCC100M: The New Data in Multimedia Research* Thomee B, Shamma DA, Friedland G, Elizalde B, Ni K, Poland D, Borth D, Li L. Communications of the ACM (2016), 59(2) 64-73
- [2] <https://www.ibm.com/blogs/research/2019/01/diversity-in-faces/>
- [3] *IBM Research Releases 'Diversity in Faces' Dataset to Advance Study of Fairness in Facial Recognition Systems* Smith JR. Accessed 2019-12-20
<https://www.nbcnews.com/tech/internet/facial-recognition-s-dirty-little-secret-millions-online-photos-scraped-n981921>
- [4] *Row over AI that 'identifies gay faces'* Accessed 2019-12-20
<https://www.bbc.com/news/technology-41188560>
- [5] *Google sells the future, powered by your personal data* Popken B. Accessed 2019-12-20
<https://www.nbcnews.com/tech/tech-news/google-sells-future-powered-your-personal-data-n870501>
- [6] *The Cost of Reading Privacy Policies* McDonald AM, Cranor LF. I/S: A Journal of Law and Policy for the Information Society (2008) <http://lorrie.cranor.org/pubs/readingPolicyCost-authorDraft.pdf>
- [7] *How much is social media worth? Estimating the value of Facebook by paying users to stop using it* Corrigan JR, Alhabash S, Rousu M, Cash SB. PLOS ONE (2018), 13(12): e0207101
- [8] *Owning Ethics: Corporate Logics, Silicon Valley, and the Institutionalization of Ethics* Metcalf J, Moss E, boyd d. Accessed 2019-12-20
<https://datasociety.net/output/owning-ethics-corporate-logics-silicon-valley-and-the-institutionalization-of-ethics/>
- [9] *Letter on Proposed Changes to the Common Rule*. Metcalf, J. 2019. Council for Big Data, Ethics, and Society. Accessed 2019-12-20.
(<https://bdes.datasociety.net/council-output/letter-on-proposed-changes-to-the-common-rule/>)
- [10] *Reading Level and Comprehension of Research Consent Forms: An Integrative Review*. Foe G, Larson EL. J Empir Res Hum Res Ethics (2016) ;11(1):31-46.
- [11] *Improving biobank consent comprehension: a national randomized survey to assess the effect of a simplified form and review/retest intervention* Beskow LM, Lin L, Dombeck CB, Gao E, Weinfurt KP. Genet Med (2017) 19(5): 505–512.
- [12] <https://gdpr-info.eu/recitals/no-32/> Accessed 2019-12-20
- [13] *Genomics, Big Data, and Broad Consent: a New Ethics Frontier for Prevention Science* Fisher CB, Layman DM. Prev Sci. (2018) 19(7): 871–879.
- [14] *Opinion: Seeking justice in the age of genomics and a call for higher ethical standards for research involving human populations* Ferguson B. Accessed 2019-12-20
<https://www.broadinstitute.org/blog/opinion-seeking-justice-age-genomics-and-call-higher-ethical-standards-research-involving-human>
- [15] *Personal Data Stores, Exchanges, and Applications* Andrieu J. Accessed 2019-12-20
<http://blog.joeandrieu.com/2010/09/17/personal-data-stores-exchanges-and-applications/>
- [16] <https://datacoup.com/docs#pp> Accessed 2019-12-20
- [17] <https://datacoup.com/docs> Accessed 2019-12-20
- [18] <https://www.crunchbase.com/organization/datacoup> Accessed 2019-12-20
- [19] *How much is your data worth to tech companies? Turns out, it's hard to say* Accessed 2019-12-20
<https://eu.usatoday.com/story/tech/2019/07/14/facebook-twitter-how-much-data-worth/39677311/>
- [20] *A tiny startup wants to pay you for your DNA, and it could lead to the next wave of medical innovation* Tracer, Z. Accessed 2019-12-20
<http://www.businessinsider.fr/us/lunadna-pays-for-genetic-information-2018-12>
- [21] *Surprise! With \$60 Million Genentech Deal, 23andMe Has A Business Plan* Harper M. Accessed 2019-12-20
<https://www.forbes.com/sites/matthewherper/2015/01/06/surprise-with-60-million-genentech-deal-23andme-has-a-business-plan/#213973052be9>

- [22] *Google will always do evil* Conditt J. Accessed 2019-12-20
<https://www.engadget.com/2018/05/24/google-will-always-do-evil/?guccounter=2>
- [23] *Trust in Facebook has dropped by 66 percent since the Cambridge Analytica scandal.* Weisbaum H. Accessed 2019-12-20
<https://www.nbcnews.com/business/consumer/trust-facebook-has-dropped-51-percent-cambridge-analytica-scandal-n867011>
- [24] *Halt the use of facial-recognition technology until it is regulated* Crawford K. *Nature* (2019) 572, 565
- [25] *France bans publishing of judicial analytics and prompts criminal penalty* Tashea J. Accessed 2019-12-20
<http://www.abajournal.com/news/article/france-bans-and-creates-criminal-penalty-for-judicial-analytics>
- [26] *France Kicks Data Scientists Out of Its Courts* Livermore M, Rockmore D. Accessed 2019-12-20
<https://slate.com/technology/2019/06/france-has-banned-judicial-analytics-to-analyze-the-courts.html>
- [27] *Refugee Roulette: Disparities in Asylum Adjudication.* Ramji-Nogales J, Schoenholtz AI, Schrag PG. *Stanford Law Review* (2007) 60(2)295-411