End of previous Forum article

Bruno Carballa Smichowski

Alternative Data Governance Models: Moving Beyond One-Size-Fits-All Solutions

In the past few years, many political and societal issues have arisen around data. The Cambridge Analytical scandal has illustrated the problems posed to both privacy and democracy by having a for-profit private firm controlling the use of detailed personal data of millions of people. Antitrust scholars and regulators such as the European Commission warn and have started taking action against anti-competitive uses of data. In many cities, transportation authorities lack access to data on ride-hailing trips or real-time traffic that is vital to their mission. As a result, companies like Uber and Waze have started selling their data to public actors. In this context, one of the main challenges the data economy faces today is the insufficient level of data sharing between public and private actors.[1]

Although these issues are heterogeneous and demand diverse policy answers, they have one common root: they all originate in what we will hereafter call the 'hegemonic data governance model'. This data governance model relies on a data collector (e.g. a platform) retaining exclusive control over the data it collects, typically through draconian clickwrap general conditions of use

Bruno Carballa Smichowski, Chronos, Paris; CEPN-Université Paris XIII, France.

1 P. Hofheinz, D. Osimo: Making Europe a data economy: a new framework for free movement of data in the digital age, in: Lisbon Council Policy Brief, Vol. 11, No. 1, 2017.

(GCU), particularly when the data collection involves individuals. After the data is collected, given that there is no such thing as *de jure* data ownership, the data collector 'owns' it de facto, although there are legal ways to protect third parties from accessing the data (yet not the data itself) through copyright over the database and/or the software that allows access to it.[2] Parallel to specific policy solutions that have been put forward to tackle each of these issues separately (the General Data Protection Regulation (GDPR), antitrust investigations, sector-specific regulations, etc.), some authors and politicians have proposed dismantling hegemonic data governance and replacing it with an alternative one. Among the most popular alternatives, two polar options have gained in popularity: either the state would make data a public good, or it could create property rights over personal and non-personal data so that a frictionless data market in which each natural or legal person can sell 'its' data can emerge.[3]

However, no one-size-fits-all alternative data governance model can respond at once to the many issues the data economy poses. This is evidenced by both existing and envisaged alternative data governance models, for which I will provide some guidelines on the scenarios and the conditions under which they might represent an alternative to the hegemonic model. Thereby I focus on the purpose of the model (what issue it tackles), the type of data it fits and its legal, technical and economic conditions of success. In particular, I will briefly examine four models: (1) crowdsourced data commons, (2) data requisition, (3) collective bargaining on rights over personal data and (4) data pooling between organisations. Data as a public good managed by the state, as we will see, is one possibility comprised in the data requisition model. I conclude by pointing out how these models can be combined to build the data economy into a variety of data governance models.

Crowdsourced data commons

Crowdsourced data commons is a governance model in which individuals collect data and pool it together to produce databases and eventually products or services based on this data that are managed as a common. A common can be defined as i) a group of self-organised actors that have set out the rules under which they intend to operate ('resource'); ii) they allocate to the various actors a set of rights and obligations regarding the way in which the pooled resource shall be treated and the benefits that may be derived and shared ('rules'); and iii) they establish forms of governance to promote the compliance with these rights and obligations ('governance').[4] Examples of this model include collaborative cartography of spaces (Open Street Maps) or transportation networks (Digital Matatus, Transport for Cairo, Jungle Bus), Wikidata, health data cooperatives such as MiData, Salus or Moipatient and several citizen science projects such as Making Sense. These initiatives typically rally a group of individuals to generate data that either did not exist before or is inaccessible in order to fulfil a societal goal (open access to fundamental data, e.g. cartographical data, advancing or orienting medical research, measuring and exposing the existence of noise or air pollution) that private and public actors do not fully tackle. In most cases, the data is open, although in some such as health data cooperatives it is not the case, as personal data is involved. The community can organise and establish the governance and the rules either informally or through a legal person such as a foundation.

This model is of particular interest in two types of scenarios that can overlap. In the first scenario, data can be more efficiently produced if crowdsourced once a critical mass of contributors is reached. It is the case of collaborative cartography: a map can be produced and updated more frequently, accurately and economically if individuals *in situ* crowdsource it. Moreover, since these projects open the data they collect, they create positive externalities for actors that can use it to produce additional services such as a route planning software. In the second scenario, crowdsourcing the data becomes the pillar of collective action. For example, pooling the data of hundreds of individuals with rare diseases and donating it to researchers allows for the advancement

2 J.E. Cohen: Law for the platform economy, U.C. Davis L. Rev. 133-2014, 2017; N. Duch-Brown, B. Martens, F. Mueller-Langer: The economics of ownership, access and trade in digital data, JRC Digital Economy Paper 2017-01, 2017.

3 M. Mazzucato: Let's make private data into a public good, MIT Technology Review website, 27 June 2018, available at https://www.technologyreview.com/s/611489/lets-make-private-data-into-a-public-good/; I. Arrieta-Ibarra, L. Goff, D. Jiménez-Hernández, J. Lanier, E.G. Weyl: Should We Treat Data as Labor? Moving beyond "Free", aea Papers and Proceedings, No. 108m, 2018, pp. 38-42; G. Koenig: Ne donnons plus nos données, Le Nouveau Magazine Littéraire, Vol. 4, 2018, p. 49.

4 B. Coriat: Le retour des communs: & la crise de l'idéologie propriétaire, Éditions Les Liens qui libèrent, 2015; P. Abecassis, J.-F. Alesandrini, B. Coriat, N. Coutinet, S. Leyronas: DNDi, a Distinctive Illustration of Commons in the Area of Public Health, 2019; F. Orsi, J. Rochfeld, M. Cornu-Volatron: Dictionnaire des biens communs, 2017, Presses universitaires de France.

in finding cures in a manner that could not have taken place without this crowdsourced data common.[5]

Crowdsourced data commons are not easy to set up and to maintain over time. We should first point out the fact that not every type of data can be subject to this governance model. The data has to be either crowdsourceable or personal data that individuals can legally claim. For example, this model does not fit industrial connected devices' data. This brings us to the model's conditions of success. First, in order for individuals to be able to crowdsource the data, they need user-friendly software to collect it, treat it and manage it. Using this software, in turn, requires training individuals to assure the accuracy and coverage of the data produced. For example, in order to collect data on noise pollution, twenty residents of the Plaça del Sol neighbourhood in Barcelona who participated in the Making Sense project went through several days of technical training on how to use and maintain the sensors they installed in their homes. Second, when the data is not produced by the individuals themselves, they need to have the legal right to claim it from the data holder in the first place. In the case of data cooperatives, this is possible because of the special legislation that exists in many countries regarding health data. In other fields, however, personal data portability is required. So far, only the European Union has created a data portability right through GDPR, although doubts remain about whether only volunteered data or also inferred and observed data are covered by this legislation.[6] Moreover, GDPR does not allow for collective data portability (i.e. many individuals deciding together to migrate personal data that links them together), which excludes relational data. Finally, the main challenge to crowdsourced data commons today is finding a sustainable business model, as they all rely on pro bono work and research grants or donations. Although cases such as Wikipedia and many alternative media sources have shown that a donations-based model can be sustained, most crowdsourced data commons struggle to assure financial stability over time.

Collective bargaining on rights over personal data

The collective bargaining on rights over personal data governance model has been prefigured by Lionel Maurel and Laura Aufrère in their article "Pour une protection sociale des données personnelles".[7] The concept of such a data governance model starts from a double observation. On one hand, there is an unbalanced power relationship between platforms unilaterally deciding the GCU that allow them to de facto own the data generated within it. On the other hand, echoing the digital labour literature, the authors observe that when using multiple platforms, individuals lose control over three things: the perception of their digital traces, the social production process of data and the use or exploitation of it in the form of explicit expression of individuals' identities.[8] This gives rise to a 'presumption of use subordination' that echoes the 'presumption of subordination' that justifies the existence of social welfare in the context of the employment contract. However, contrary to traditional labour relationships framed in an employment contract, digital labour comprises a succession of practices that results in a continuum of statuses ranging from users (e.g. a Facebook user producing data as a by-product of a leisure activity) to platform workers (e.g. an Uber driver whose professional data is exploited). As the value of this socially produced personal data lies in its relational nature,[9] there are mechanisms inspired by those of social welfare that would allow people to collectively bargain over the rights to their personal data that could be justifiable and virtuous.[10]

In practice, this would translate into legally recognised collective entities such as unions that would bargain over the GCU. It is important to point out that these entities would differ from the 'data unions' imagined by authors that defend property rights over data,[11] as the social welfare approach excludes the possibility of creating such rights. Data unions could be both sector-specific (e.g. social networks, crowdsourced review

5   A. Blasimme, E. Vayena, E. Hafen: Democratizing Health Research Through Data Cooperatives, in: Philosophy & Technology, Vol. 31, No. 3, 2018, pp. 473-479; E. Hafen, D. Kossmann, A. Brand: Health data cooperatives – citizen empowerment, in: Methods of Information in Medicine, Vol. 53, No. 02, 2014, pp. 82-86.
6   J. Crémer, Y.-A. Montjoye, H. Schweitzer: Competition policy for the digital era, Directorate-General for Competition, European Commission, 2019.
7   L. Maurel, L. Aufrère: Pour une protection sociale des données personnelles, SI Lex, 5 February 2018, available at https://scinfolex.com/2018/02/05/pour-une-protection-sociale-des-donnees-personnelles.
8   D. Cardon, A. Casilli: Qu'est-ce le digital labor?, Bry-sur-Marne, INA, coll. "Etudes et controverses", 2015; C. Fuchs: Digital Labor, The Routledge Companion to Labor and Media, p. 51, 2015.
9   For example, it is the network of relations between individuals' digital footprints, and not each individual's data, which is valuable to a platform like Facebook.
10 Let us make clear that the term 'social welfare' is used in a wide manner. It is understood as a macro-system of social, legal and political relations between the domestic economic and political spheres that protect individuals and their families to live in dignity against life hazards and society against disintegration forces threatening it, see N. Alix, L. Aufrère, J.-C. Barbier, J.-C. Boual, F. Hermet, S. De Heusch, H. Vandenbilcke: La protection sociale en France: une macro institution en réforme permanente, Perspectives du point de vue de l'ESS et des commun. Groupe de recherche collaborative protection sociale, ESS et Communs au sein de la Coop des Communs, 2018.
11 I. Arrieta-Ibarra et al., op. cit.

websites, ride-hailing apps, etc.) or territorially based (a city, a region, a country). They could bargain over measures to collectively protect their privacy or the sharing of data with third parties. For example, a lodging platform's data union could bargain over the donation of data produced by platform users to government bodies that would use it for regulation and urban planning purposes. It is interesting to point out that burgeoning versions of this data governance model are starting to emerge (in which users act collectively to level the playing field against platforms). For example, there have been class action lawsuits against Facebook concerning the misuse of users' personal data.[12] In platforms such as Uber or Deliveroo, in turn, drivers have been known to coordinate to log off simultaneously so that the price of a ride or delivery would rise automatically.

This governance model can only apply by definition to relational personal data massively collected by platforms such as Facebook, Yelp, Uber or Airbnb. In order to exist, it would require a major evolution that incorporates the necessary legal mechanisms to establish data unions and social rights over socially produced personal data. Moreover, contrary to traditional unions, most data unions would not have the means to go on strike in order to pressure platforms. Indeed, even after big scandals, users have not stopped using platforms en masse, as the platforms are part of users' everyday life (contrary to traditional workspaces like a factory or an office) and they are the medium through which many users that do not know each other personally relate to each other. Therefore, legal mechanisms such as an institution that could force platforms to negotiate with data unions and workable user-friendly open-source privacy-aware alternative platforms based on data commons would be needed to render data unions' collective bargaining feasible. This opens up many questions but I focus on two: collective portability of personal data (to enable users to migrate easily to alternative platforms) and the sustainability of alternative platforms and data commons' business models. The latter is of particular relevance due to the fact that many hegemonic content-based platforms' business models rely on targeted advertising and hence on invasive personal data harvesting.

Data requisition

'Data requisition' is a term used to describe a variety of situations in which a public actor demands that a private actor shares data either in exchange for payment or for free. The sharing could take three scopes, each defining a sub-model. It could be shared only with public actors for regulation purposes. This is what the French Member of Parliament Luc Belot intended when he proposed creating the legal concept of 'territorial interest data' which would give regional governments that exact power.[13] The public actor could also demand that a private actor opens certain datasets. Again, France is a pioneer country. The *Loi pour une Répulique Numérique* (Law for a Digital Republic) created in 2016 gives municipalities the right to demand private operators to open their data if

1. the private actor benefits from a public service delegation contract;
2. the private actor's activity relies on at least a certain threshold of public subsidies;
3. the private actor holds energy consumption data generated by public infrastructure or
4. in certain cases, if there is jurisprudence data involved.

Moreover, the French Parliament approved in June 2019 the *Loi d'Orientation des Mobilités* (Law of Orientation of Mobilities, LOM), which includes articles that will force mobility operators and some platforms to open static and real-time user information data (location of a bus, delays, location of free-floating scooters, etc.). Finally, the public actor could allow a private actor to share some datasets on fair, reasonable and non-discriminatory terms with certain third parties, typically to avoid anti-competitive refusals to grant access to data, as suggested by the European Commission in its recent report 'Competition policy for the digital era'.[14]

Data requisition is of particular interest in cases in which private actors do not have incentives to share data and the reluctance to share it has a negative impact on public interest. For example, the data may be needed by the public actor to fulfil its mission of guarantor of the public interest (e.g. regulating a market, producing public statistics). Under certain circumstances, the lack of sharing could also harm competition if the data hoarded by a firm is difficult to reproduce or bypass and nonetheless necessary for other firms to produce a similar product or service. Finally, making private data a public good by forcing private actors to open it can create positive externalities that justify the requisition. For example, in the

12 K. Mehrotra, A. White: Facebook Must Face Lawsuit Over 29 Million-User Data Breach, Bloomberg, available at https://www.bloomberg.com/news/articles/2019-06-24/facebook-must-face-lawsuit-over-29-million-user-data-breach.

13 L. Belot: De la Smart City au territoire d'intelligence (s)–L'avenir de la Smart City, rapport au Premier ministre sur l'avenir des smart cities, 2017.
14 J. Crémer et al., op. cit.

case of the LOM, this will not only improve regulators' knowledge of the mobility sector, but it will also allow for the development of a variety of data-driven services that might not arise otherwise, notably mobility-as-a-service (MaaS) platforms. Indeed, transport operators fearing that the data sharing needed to develop a MaaS offering will result in losing clients to other transport operators or aggregators would back off, which would delay the development and the quality of MaaS platforms.[15]

Although powerful, the data requisition model requires major legislation changes or the application of existing legislation to data. While the former can be politically difficult to achieve, the latter can bring difficulties. For example, as pointed out by Crémer et al.,[16] applying the essential facility doctrine to data,[17] which had been developed to target physical infrastructures, presents problems. Moreover, when adapting or creating legislation, one should take into account the avoidance of discouraging private actors' generation of data. This leaves data constituting trade secrets or a core competitive advantage for a firm out of the scope of data requisition.

Data pooling between organisations

Data pooling between organisations is the most common alternative data governance model. When organisations, both public and private, hold complementary data (i.e. the overall utility and/or exchange value of the datasets that each agent holds is increased if combined), they have an incentive to create data pools. We can distinguish two sub-models in terms of the bundle of rights they create around data: open data pooling and closed data pooling.

In the case of open data pooling between organisations, non-profit-oriented organisations usually pool data to better fulfil their non-commercial mission. For example, the regional governments of Bretagne and Pays de Loire in France contribute to an open data pool called PRIDE in order to design better energy policy using the more accurate and exhaustive information that comes from an enlarged and enriched dataset allowed for by open data. Another example of this logic is Transdev's Catalogue, a platform of pooled open transportation

data. Profit-oriented firms, in turn, have many commercial motivations, notably creating a related business, good publicity (when their opened data helps to tackle a societal issue), gaining expertise and increasing interoperability. In the case of closed-data pooling between organisations, non-profit-oriented organisations have the same motivation to better fulfil their non-commercial mission, but they cannot choose an open data-pooling scheme for several reasons. The most common reasons are privacy protection (when personal data is involved), security (e.g. sensitive data on water and energy infrastructures) and economic risk, typically when private stakeholders participating in the data pooling are reluctant to open data that has a commercial or strategic value and their business models are not compatible with an open data strategy.[18]

An example of closed data pooling are integrated digital care records. In several territories of the United Kingdom, public and private care sector actors pool data in order to create integrated digital care records that facilitate patients' treatment and the production of accurate and holistic health statistics at the territorial level.[19] Given the sensitivity of the personal data involved, the data has not been opened. Profit-oriented organisations, in turn, revert to closed data pooling for four main reasons. First, data can be pooled to improve the collaboration and therefore the joint value creation within a supply chain or ecosystem. This is the case with the Airbus Skywise platform, a data-sharing platform set up by Airbus to pool data between the firms that make up the value chain (OMEs, airlines, maintenance companies, etc.) and provide predictive analysis based on this data to the fluidity of the work between them. Second, profit-oriented organisations may have incentives to pool data to build a new product or service. The best example is MaaS platforms such as Whim or Compte Mobilité, which can only exist if several transportation operators and other mobility actors, such as firms running parking spaces, share data with each other. Finally, mirroring certain patent pools, profit-oriented firms can share closed data aimed at foreclosing competition,

15 B. Carballa Smichowski: Determinants of coopetition through data sharing in MaaS, in: Management & Data Science, Vol. 2, No. 3, 2018.

16 J. Crémer et al., op. cit.

17 The following have already envisioned the possibility of applying the essential facility doctrine to data: Z. Abrahamson: Essential data, in: Yale Law Journal, Vol. 124, No. 3, 2014, p. 867; and I. Graef: Data as essential facility: competition and innovation on online platforms, 2016.

18 B. Carballa Smichowski: The value of data: an analysis of closed-urban-data-based and open-data-based business models, Working paper No. 01/2018 of the Cities and Digital Technologies Chair, Urban School, Sciences Po, 2018; S. Chignard, L.-D. Benyayer: Datanomics. Les nouveaux business models des données, FYP editions, 2015.

19 For a case-study-based analysis of data governance in integrated digital care records see Future Care Capital: Intelligent sharing: unleashing the potential of health and care data in the UK to transform outcomes, 2017.

which poses questions in terms of antitrust application to the digital economy.[20]

It is interesting to point out that both open and closed data pooling are usually initiated by an actor that plays the role of the 'orchestra conductor' i.e. rallying other organisations. This role is typically played by an actor that is legitimate and/or has a tighter relationship with the other organisations pooling data (e.g. a municipality in a city level data sharing experiment) or by the strongest actor among the poolers, as is the case of Airbus in the Airbus Skywise platform.

Although promising, this data governance model faces many obstacles even when (in principle) organisations are open to the idea of pooling their data. In economic terms, the main obstacle is a firm's fear that the data pooling offers its partners a competitive advantage that will be used against them. MaaS platforms are a good example of this phenomena: by pooling data, transportation operators engage in a cooperative dynamic in which they can increase the overall customer base, but they might also lose clients from partners-competitors and be expelled from the market by the aggregator, who retains the customer relationship.[21] In technical terms, the main obstacle is the absence of a technical standard to share data, which makes it difficult and costly.[22]

Moreover, in terms of applicability, it should be noted that data pooling between organisations is facilitated when certain conditions are met. First, as mentioned above, there must be complementarity between the data held by the different organisations. This means that the combined dataset has to have more value (either in itself or through a service or product it allows to create or improve) than the sum of the value of the separate datasets. However, for all the parties to have incentives to create this joint value by pooling data together (as opposed to simply selling the data or refusing to share it), this value has to be symmetric in the sense that all the parties should need each other's data to create more value. Second, in the case of data pooling between actors competing in the same market, data sharing tends to occur when market concentration is low. In that case, no actor is big enough to do without other actors' data, as in the case of mobility. Third, data pooling is more

likely to take place when the technical conditions of data production allow many actors to produce data that, when pooled together, can create more value. A counterexample is energy transportation and distribution, which can only be produced by the infrastructure manager, thereby making data pooling less necessary.

Conclusion

Data governance is currently dominated by the hegemonic model in which the data collector retains exclusive control over the data it collects. The overreach of this model has created problems in various fields that call for alternative data governance models. I have studied four emerging or envisioned alternative data governance models outside of the more known public open data: crowdsourced data commons, collective bargaining on rights over data, data requisition and data pooling between organisations. For each of them, I have examined the conditions of applicability and the main legal, economic and technical obstacles they face. This leads to my main conclusion: there is no one-size-fits-all alternative data governance model. Irrespective of regulators and actors' objectives, a workable alternative data ecosystem can only be built on a variety of data governance models.

This conclusion shifts our point of analysis from data governance models to a data ecosystem made of several data governance models. This shift and the previous analysis of the functioning, potential and limitations of each model, allows us to reach a second conclusion that deserves further research: alternative data governance models can complement each other. Indeed, by forcing an organisation to share its data with other ones or even opening it, the data requisition model can feed crowdsourced data commons models and trigger data pooling between organisations. When the public actor forces other actors to open their data, this also feeds the public open data model, which in turn feeds the entire ecosystem – including firms that use this open data although they govern their own data with the hegemonic model. Crowdsourced data commons, in turn, can create a legal entity that integrates a data-pooling scheme between organisations, but they can also feed the hegemonic data governance model by generating data that traditional firms can use to develop a data-driven service. Finally, collective bargaining on rights over personal data can function as a counterweight in terms of the defence of a collective's fundamental rights when their data is being governed by an actor recurring to the hegemonic model or by a group of organisations pooling data.

20 P.-A. Mangolte: La guerre des brevets d´Edison aux frères Wright: Une comparaison franco-américaine, Paris 2014, Éditions l´Harmattan.
21 B. Carballa Smichowski: Determinants of coopetition through data sharing in MaaS, op. cit.
22 C. Arnaut, M. Pont, E. Scaria, A. Berghmans, S. Leconte: Study on data sharing between companies in Europe, DG Communications Networks, Content & Technology, 2018, European Commission.