Data collaboratives, competition law and the governance of EU data spaces

Nicolo Zingales*

[Forthcoming in I. Kokkoris (ed.), *Research Handbook in Competition Enforcement* (Edward Elgar)]

Introduction

We live in an era of profound digital transformation, and this has implications on the way in which businesses bring their products to markets and conceive of those products in the first place. In an increasingly connected economy, data (be it of personal or non-personal nature) represent a fundamental driver of competition and innovation, providing a potentially insurmountable lead advantage[1] and an ability to leverage customer relationship across multiple products and markets[2]. And while EU consumers, professional users, and data subjects enjoy a right to data portability in various forms[3], the effectiveness of this right is questionable from the perspective of creating a level playing field, particularly where data is used to consolidate a leadership position within and across markets. Against this background, a number of studies have suggested the imposition of data access and data sharing obligations[4]. Some of these obligations have been included in the proposed Digital Markets Act and Digital Services Act[5], which either presuppose or encourage the creation of repositories of information

* Professor at FGV Law School, Rio de Janeiro. Comments welcome at nicolo.zingales@fgv.br

[1] Prufer, Jens and Schottmüller, Christoph, Competing with Big Data (February 16, 2017). TILEC Discussion Paper No. 2017-006, CentER Discussion Paper 2017-007, Available at SSRN: https://ssrn.com/abstract=2918726 or http://dx.doi.org/10.2139/ssrn.2918726

[2] Padilla, Jorge and Daniele Condorelli, "Data-driven Envelopment with Privacy-Policy Tying" (2020) <papers.ssrn.com/sol3/papers.cfm?abstract_id=3600725> accessed 10 January 2021

[3] Inge Graef, Nadezdha Purtova and Martin Husovec, "Data portability and data control: Lessons for an emerging concept in EU law" (German Law Journal, 19(6), 1359-1398, 2018)

[4] OECD, Enhancing Access to and Sharing of Data: Reconciling Risks and Benefits for Data Re-use across Societies (2019). Jens Prüfer. Competition Policy and Data Sharing on Data-driven Markets: Steps Towards Legal Implementation. Friedrich Ebert Stiftung (2020). Graef, Inge and Tombal, Thomas and de Streel, Alexandre, Limits and Enablers of Data Sharing. An Analytical Framework for EU Competition, Data Protection and Consumer Law (November 27, 2019). TILEC Discussion Paper No. DP 2019-024, Available at SSRN: https://ssrn.com/abstract=3494212 or http://dx.doi.org/10.2139/ssrn.3494212; Teresa Scassa, "Sharing Data in the Platform Economy: A Public Interest Argument For Access to Platform Data", 10 (4) UBC law Review 1017; Teresa Scassa, Designing Data Governance for Data Sharing: Lessons from Sidewalk Toronto, Technology and Regulation, 2020, 44–56 https://doi.org/10.26116/techreg.2020.005. Vikas Kathuria and Jure Globocnik, Exclusionary conduct in data-driven markets: limitations of data sharing remedy, Journal of Antitrust Enforcement. jnz036, https://doi.org/10.1093/jaenfo/jnz036; Heiko Richter, Exposing the Public Interest Dimension of the Digital Single Market: Public Undertakings as a Model for Regulating Data Sharing (March 21, 2020). Max Planck Institute for Innovation & Competition Research Paper No. 20-03, Available at SSRN: https://ssrn.com/abstract=3565762 or http://dx.doi.org/10.2139/ssrn.3565762

[5] Proposal for a Regulation of the European Parliament and the Council on contestable and fair markets in the digital sector (Digital Markets Act) COM/2020/842 final; and Proposal for a Regulation of the European Parliament and the Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC COM/2020/825 final. In the Digital Markets Act (DMA), see art. 6, which requires online gatekeepers to "(i) provide business users, or third parties authorised by a business user, free of charge, with effective, high-quality, continuous and real-time access and use of aggregated or non-aggregated data, that is provided for or generated in the context of the use of the relevant core platform services by those business users and the end users engaging with the products or services provided by those business users; for personal data, provide access and use only where directly connected with the use effectuated by the end user in respect of the products or services offered by the relevant business user through the relevant core platform service, and when the end user opts in to such sharing with a consent in the sense of the Regulation (EU) 2016/679; and

accessible to qualified actors. In parallel, the European Commission is promoting for strategic economic sectors and domains of public interest the concept of "common European data spaces", defined as "large pools of data in these sectors and domains, combined with the technical tools and infrastructures necessary to use and exchange data, as well as appropriate governance mechanisms"[6]. However, we are yet to know the rules, procedures and institutional safeguards applicable to these data sharing mechanisms, which are undoubtedly crucial when it comes to determining their legality under competition law. This chapter sketches the challenges involved in this determination for a wide range of collaborations revolving around data access and data sharing, which we call "data collaboratives".

The chapter is structured as follows: section 1 provides an introduction to the notion of data collaboratives and their scope, followed by an illustration of the peculiarities of its different forms: data cooperatives, data unions, data trusts, data pools and data exchanges. Section 2 investigates where the concept of "EU data space" may fit along this spectrum, based on the information disclosed in the Commission's data strategy. Section 3 focuses on the way in which the current competition law framework in EU views these collaborations, with particular focus on the challenges raised by the intrinsic economic value of data and its role of input for entering new markets. Section 4 briefly concludes, summarizing the learning from this exercise and suggest a way in which regulators can improve legal certainty through the use of innovation hubs and regulatory sandboxes.

1. The rise of data collaboratives

To begin discussing the promises and challenges for competition law in this domain, we must first of all understand the phenomenon under examination. In this regard, an important preliminary question for our inquiry concerns its scope: this is a jurisdictional question, because competition law does not apply indistinctively to all human activities, in every sector of the economy; but also, a substantive question, as the legal and economic context of collaborations has great relevance on the application of competition law.

We have chosen here a term that goes beyond the more restricted notion of bottom-up initiatives to pool or exchange data[7] in order to reflect the fact that these "initiatives" may simply be the result of regulatory obligations, thus drawing attention to the different set of considerations that apply when these partnerships operate against the backdrop of the sanctioning power of the State. At the same time, this broad categorization of "data collaboratives" goes beyond the original use of the term coined by researchers at the NYU Governance Lab who referred to "a new form of collaboration, beyond the public-private partnership model, in which participants from different sectors — including private companies,

(j) provide to any third-party providers of online search engines, upon their request, with access on fair, reasonable and non-discriminatory terms to ranking, query, click and view data in relation to free and paid search generated by end users on online search engines of the gatekeeper, subject to anonymisation for the query, click and view data that constitutes personal data.

In the Digital Services Act (DSA), see art. 31, which imposes on very large online platforms the obligation to grant access to data to "vetted researchers […] for the sole purpose of conducting research that contributes to the identification and understanding of systemic risks […]"; and art 34 (1) (f) which refers to EU Commission's promotion of voluntary standards for "transmission of data between advertising intermediaries in support of transparency obligations pursuant to points (b) and (c) of Article 24".

[6] Communication From the Commission to the European Parliament, The Council, The European Economic and Social Committee and the Committee of the Regions, A European Strategy for data (February 2020) <https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1593073685620&uri=CELEX%3A52020DC0066>, p. 22

[7] a category where we can fit data pools, data cooperatives and data exchanges, see below in section 1.

research institutions, and government agencies — can exchange data to help solve public problems"[8]. This term has acquired in the data governance literature the connotation of collaboration which is aimed to address public policy issues; whereas the present analysis includes initiatives that are purely profit-driven, and that may be designed to overcome problems of *private* (rather than public) ordering. Having said that, it should also be recognized that these initiatives typically arise because of the existence of a perceived market failure: i.e., a negative externality in the operation of the market mechanism that is not adequately addressed, either through public or private regulation. In this sense, it is important not to underestimate the significance of the efficiencies generated by these initiatives, particularly when they fill gaps left by regulators that struggle to keep up with the swirling advancement of digital technologies and their effects on society.

The increasing impact of datafication on our economy presents us with new ways of addressing these market failures, but also come with different types of associated harms, making it necessary to further develop existing regulatory approaches. This includes a reconsideration of domains where we might want to encourage collaboration rather than competition because the latter is toxic, or socially detrimental when compared to its benefits[9]. Depending on the distribution of harms, benefits and risks involved, it might be preferrable to configure different degrees of interactions between markets participants, especially when it comes to the roles and responsibilities with regard to information sharing. Some of these collaborations may take place under the framework of well-recognized legal structures, while others may evade the boundaries of these classifications and give rise to new forms of co-creation. A simple ontology is attempted below, focusing on the degree and forms of integration of the activities of participants.

Data cooperatives

Cooperatives are the most logical departure point for our analysis, as they involve integration through common ownership and control. Broadly speaking, a cooperative is "an autonomous association of persons united voluntarily to meet their common economic, social, and cultural needs and aspirations through a jointly-owned and democratically-owned enterprise"[10]. Thus, in addition to the peculiar ownership structure, which ensures its independence from profit-driven financial investors, cooperatives can be distinguished for what has been called "the essential characteristic of a cooperative": its fundamentally democratic nature, which makes it impossible for its management to escape the control of its members[11]. Cooperatives are

[8] Stefaan Verhulst and David Sangokoya, "Data Collaboratives: Exchanging Data to Improve People's Lives" (The GovLab, 22 April 2015) <https://sverhulst.medium.com/data-collaboratives-exchanging-data-to-improve-people-s-lives-d0fcfc1bdd9a> accessed 10 January 2021

[9] See in this sense, specifically referring to competition among social media: Nicolas Petit, *Big Tech and the Digital Economy* (Oxford University Press 2020), p. 243. See more generally: Ariel Ezrachi and Maurice E. Stucke, *Competition Overdose: How Free Market Mythology Transformed Us from Citizen Kings to Market Servants* (Harper Business 2020); Michelle Monger, *Competition is Killing Us: How Big Business is Harming Our Society and Planet - and What To Do About It* (Penguin Books 2020);

[10] See the Statement on the Co-operative Identity" adopted by the International Co- operative Alliance (ICA) in 1995, endorsed in UN resolution 56/114 adopted at the 88th Plenary meeting of the U.N. General Assembly on 19th December 200; UN Report of the Secretary General 2001/68 dated 14th May 2001; and Recommendation 193 on the promotion of Cooperatives adopted at the 90th Session of the International Labour Conference on 20th June 2002, Section I. 2.

[11] See Ian McPherson, "The Co-operative Identity in the Twenty First Century", Review of International Co-operation 3/94, Geneva, 1994: "The essential characteristic of a cooperative is that it is a democratic organisation engaged in the market place, providing goods and services. It is nevertheless based on people, not on capital or

modelled upon a set of values and principles that have been enshrined in the Statement of Cooperative Identity, adopted by the International Cooperative Alliance in 1995[12].

In EU legislation, the concept of cooperative finds specific recognition with Regulation on the Statute for the European Cooperative Society of 2003 (issued together with Directive 2003/72, supplementing the Statute for a European Cooperative Society with regard to the involvement of employees) and the EU Commission's Communication the promotion of co-operative societies in Europe of 2004[13]. The Regulation, in particular, sets out in its Article 1 that a cooperative which wishes to constitute itself as a European Cooperative Society (SCE) shall have as its principal object the satisfaction of its members' needs and/or the development of their economic and social activities, in particular through the conclusion of agreements with them to supply goods or services or to execute work of the kind that the SCE carries out or commissions[14]. Its key requirements are the existence of five or more natural or legal persons resident or governed by the law of at least two EU member states[15], and holding a minimum capital of EUR 30 000[16]. According to the Regulation, cooperatives may extend the benefits of their activities to non-members or allow them to participate in its business, except where its statutes provide otherwise[17]. This means that there is a significant degree of flexibility on the institutional and organizational structure that cooperatives can take. A large discretion is left to the statutes of the cooperatives on a number of strategic issues, including how many shares are necessary to qualify for membership, with the important caveats that each member has a vote (a fundamental tenet of the principle of democratic control)[18] , and that *if* the statutes stipulate that the majority at general meetings shall be constituted by members who are natural persons and there is a subscription requirement for members wishing to take part in the activities of the SCE, the statute may not make membership subject to subscription for more than one share[19]. More complex rules and exceptions can be found at the national level for the constitution of cooperatives[20]; however, this contribution remains focused on the EU dimension in light of the harmonizing role of the Regulation, and its potential in providing bright-line rules for our analysis.

government direction. In its essence it can never escape, even if it wanted to, the capacity of members to exercise control whenever they wish to do *so".*

[12] See International Cooperative Alliance, "Cooperative identity, values & principles" (ICA, 2018) <https://www.ica.coop/en/cooperatives/cooperative-identity#:~:text=The%20Statement%20on%20the%20Cooperative,and%20democratically%2Dcontrolled%20enterprise.%E2%80%9D>. Values include those of self-help, self-responsibility, democracy, equality, equity and solidarity, honesty, openness, social responsibility and caring for others; whereas the principles are those of voluntary and open membership; democratic member control; member economic participation; autonomy and independence; education, training and information; cooperation among cooperatives; and concern for community.

[13] Statute for the European Cooperative Society [2003]; Council Directive 2003/72/EC - supplementing the Statute for a European Cooperative Society with regard to the involvement of employees [2003]; Communication from the Commission to the Council and the European Parliament, the European Economic and Social Committee and the Committee of Regions on the promotion of co-operative societies in Europe [2004].

[14] Council Regulation (EC) No 1435/2003 on the Statute for a European Cooperative Society (SCE) [2003], Art 1 (1)- 1 (3).

[15] *Id*., Art 1 (5)

[16] *Id*., Art. 3 (2)

[17] *Id*., Art. 1 (4)

[18] *Id*., Art. 59 (1)

[19] *Id*., Art. 7

[20] European Parliament Research Service, Cooperatives: Characteristics, activities, status, challenges Annex < https://www.europarl.europa.eu/RegData/etudes/BRIE/2019/635541/EPRS_BRI(2019)635541_EN.pdf > accessed 01 June 2021

The 2004 Communication is a more policy-oriented instrumented, which builds on the newly established EU framework with a view to promoting wider utilization of cooperatives, in particular as vehicles for the implementation of many Community objectives in fields like employment policy, social integration, regional and rural development, agriculture, etc. Indeed, it is clear that cooperatives can be a more suitable form of partnership for the pursuit of objectives such as sustainability, social justice or more inclusive economic growth. For instance, an ILO study conducted before the financial crisis of 2008 found that cooperative financial institutions were as efficient as traditional banks, but also more stable[21]. Similarly, it has been noted that a large market share for regional cooperatives can decrease price volatility (in the agricultural sector, with regard to dairy products in particular)[22].

The goal of the Communication is to facilitate wider use of the insufficiently utilized potential of cooperatives. To that end, the Commission highlights the distinctive advantages of cooperatives in increasing economic power of SMEs, guaranteeing high quality services, and generating intangible assets of knowledge and skills through participatory management, and reflects on a number of issues that had been raised during its extensive consultation in 2002. This includes the role of competition law with regard to cooperatives, in respect of which Commission's position is that cooperatives qualify as economic undertakings and therefore are fully subject to competition law, but certain aspects of their legal form and structure warrant special consideration on a case-by-case basis. This is not elaborated upon in the text, but from the Commission's reference to earlier case-law we can extrapolate that some degree of coordination and exclusivity may be necessary to allow for the proper functioning of a cooperative[23]. Further, reference can be made to one of the key guidelines set out by ILO for the development of a supportive and coherent legal framework for cooperatives: to "encourage the development of cooperatives as autonomous and self- managed enterprises, particularly in areas where cooperatives have an important role to play or provide services that are not otherwise provided"[24].

It is natural to expect that in a data-driven economy, many of these functions can be effectively fulfilled by making strategic use of the data available to some of the participants. This is why we are seeing multiple organizations emerge to provide the architecture for collective ownership of data, through secure and trustworthy data aggregation and permissioned sharing. A leading example is MIDATA, a nonprofit cooperative based in Switzerland that operates a data platform, acts as a trustee for data collection and guarantees the sovereignty of citizens over the use of their data[25]. In addition to providing a storage on a platform which enables individuals to grant access to third parties on an individual basis, currently within the specific domain of health data and smartphone app-based services, MIDATA offers its members the

[21] Birchall, Johnston, Resilience in a downturn: The power of financial cooperatives, Geneva: ILO, 2013 Available at http://www ilo org/empent/Publications/WCMS_207768/lang--en/index htm

[22] European Parliament Research Service, Briefing. Cooperatives: Characteristics, activities, status, challenges (2019). Available at https://www.europarl.europa.eu/RegData/etudes/BRIE/2019/635541/EPRS_BRI(2019)635541_EN.pdf

[23] See for instance Case C-250/92 *Gøttrup-Klim Grovvareforening and others and Dansk Landbrugs Grovvareselskab AmbA (DLG)*, EU:C:1994:41 (holding that "A provision in the statutes of a cooperative purchasing association, forbidding its members to participate in other forms of organized cooperation which are in direct competition with it, is not caught by the prohibition in Article 85(1) of the Treaty, so long as the abovementioned provision is restricted to what is *necessary to ensure that the cooperative functions properly and maintains its contractual power in relation to producers*"). (emphasis added)

[24] ILO, Recommendation on the promotion of cooperatives, section 6 (e).

[25] See <https://www.midata.coop/en/home/> accessed 1 June 2021

option to join the cooperative as members and therefore contribute to its governance on an equal basis with other members.

Salus Coop is another domain-specific application of this approach: it is a non-profit data cooperative for health research founded in Spain in 2015, that allows individuals to donate their data for health research on an anonymized basis[26]. It works through a "common good data license" for health research', with the following conditions:

- health only: the data will only be used for biomedical research activities and health and/or social studies
- non-commercial: research projects will be promoted by entities of general interest, such as public institutions, universities and foundations
- shared results: all research results will be accessible at no cost
- maximum privacy: all data will be anonymised and unidentified before any use
- total control: members can cancel or change the conditions of access to their data at any time.

Another example is Driver Seat, a driver-owned cooperative incorporated in the USA in 2019[27]. According to the Ada Lovelace Institute[28], it allows on-demand drivers to track the data they generate, and share it with the cooperative, which can then aggregate and analyse them to produce wider insights, which members can use to optimise their incomes. Driver's Seat Cooperative also collects and sells aggregated mobility insights to city agencies to enable them to make better transportation-planning decisions. Furthermore, when 'the Driver's Seat Cooperative profits from insight sales, driver-owners receive dividends and share the wealth'.

Finally, it bears mentioning of similar examples of collective empowerment through altruistic data donation in the context of municipalities, as in the pilot projects of La Rochelle, Nantes and Lyon[29]. Specifically, Nantes has begun using data made available by citizens to develop an energy transition scheme for the city; La Rochelle intends to improve mobility services and public transport through insights gained from such data; and Lyon aims to help socially excluded families and to simplify the life of citizens who do not speak French[30].

To understand the effects of these arrangements, it is useful (see also below, section 4) to resort to the division drawn by the ILO between four different types of cooperatives based on the interests involved[31]: (a) producer cooperatives, which bring together producers of the same product or service to share assets for production, or to obtain better rates in purchasing inputs for production; (b) consumer/user cooperatives, which imply that a particular business is (at least to some extent) owned and controlled by its *customers;* (c) worker cooperatives, where businesses are owned and controlled by their employees[32]; and finally (d) multi-stakeholder

[26] Although the technological architecture allows individuals to know about and manage any data they contribute; and members always have the right to obtain openly and at no cost the results of studies carried out with the use of their data. See Salus Coop. Available at: <https://www.saluscoop.org/> accessed 01 June 2021
[27] See Driver's Seat Cooperative. Available at: <www.driversseat.co> accessed 1 June. 2021.
[28] Ada Lovelace Institute, "Exploring legal mechanisms for data stewardship" (Ada Lovelace Institute 4 March 2021) <https://www.adalovelaceinstitute.org/report/legal-mechanisms-data-stewardship/> accessed 01 June 2021
[29] Impact Assessment on the Data Governance Act [2020], p. 17
[30] CNIL (2017) La plateforme d'une ville, CAHIERS IP N.5.
[31] ILO, Guidelines concerning statistics of cooperatives, 20th International Conference of Labour Statisticians Geneva, 10-19 October 2018
[32] However, this is not to be confused with employee stock option plans, which do not confer democratic control on employees; nor is it equivalent to workers' or trade unions, which typically cater to more diffuse interests than

cooperatives, where no type of member has a dominant position through a majority of votes in the governing body or an exclusive veto over decisions. In principle, the later model is likely to be the least problematic from a competitive standpoint: the more uniform the interests of those who control the cooperative, the more the cooperative might be seen as a vehicle for collusion or for exclusionary conduct. Ultimately, the composition and governance structure of the cooperative is likely to have significant implications not only for the benefits that its members enjoy, but also for the competitive assessment of the underlying arrangements. They are also important to distinguish cooperatives from trade unions and investor-owned business, as illustrated in Table 1 below.

	Trade Unions	Cooperatives	Investor owned businesses
Purpose	Representing and defending member needs in workplace and sometimes more widely.	Meeting needs of members: economic, social, and cultural.	To generate proft for shareholders
How is it run?	Normally democratic, one member one vote. May have elected committee.	Management answers to board of democratically elected members, on one member one vote principle. Non-transferable shares.	Board answerable to shareholders. Votes determined by size of shareholding. Transferable shares.
Ownership	Owned by the members.	Owned by the members.	Publically traded shares. Price can fuctuate.
Where do profits go?	Do not generate much income beyond members' contributions – no proft.	Reinvested in cooperative or community, and/or member dividend, based on trade with and via the cooperative.	To shareholders and invested in business.
Income	Membership fees	Charges for goods and services.	Charges for goods and services.
Key advantages	Can be part of wider movement. Provides voice. Can change working conditions.	Inclusive business model. Local ownership. Generates income for members. Values based. Part of wider movement.	Focus on proft maximisation. Can raise capital via private investors, markets.
Key challenges	Find it diffcult to operate where standard employeremployee relationship does not exist.	Problems raising capital. Historical legacy of government control. Skills defcit among board members. Lack of enabling environment.	Governance challenges. Short term approaches. Proft maximisation can override ethical considerations.

Table 1. Adapted from Cooperatives Europe, *Building Strong Development Cooperation: Partnership Opportunities Between Cooperatives and the EU* (September 2015)

Data Unions

"Data union" is another important concept to facilitate the production of collective value through an organized labor movement, in particular by collecting and coordinating instances of data workers. This categorization has not been formally recognized by trade unions, as the idea of data as labor is incipient and undertheorized[33]. Nevertheless, the idea is part of a broader

those of their members (thereby also benefitting non-members) and are not supposed to generate profits. For a graphic illustration of the difference between these three, see Table 1. See also the next subsection, "Data Unions".

[33] A good introduction to this concept is provided in Arrieta Ibarra, Imanol and Goff, Leonard and Jiménez Hernández, Diego and Lanier, Jaron and Weyl, Eric Glen, Should We Treat Data as Labor? Moving Beyond 'Free' (December 27, 2017). American Economic Association Papers & Proceedings, Vol. 1, No. 1, Forthcoming, Available at SSRN: https://ssrn.com/abstract=3093683.

rising movement that seeks to promote the unionization of digital workers, which sees promising potential in crowdsourcing platforms to provide information on workers' rights, collect workers' grievances, and coordinate or mobilize collectives[34].

The goal of data unions goes even further than that, by creating stakeholder representation structures that enable negotiations between management and labor. This collective action has been successfully conducted for digital platform workers (more specifically, professional content creators) on Youtube, through a movement that enabled workers to overcome the organizational, technological and geographical fragmentation that affected their ability to engage in effective negotiation and thus improve their working conditions[35]. Since an important element of the Youtuber Union's strategy was the cooperation with a German trade union, which offered them the prospect of taking the platform in question to court[36], it is not clear to what extent these grassroot movements would be likely to impose effective constraints without support from the existing union structure. However, similarly promising signs can be seen in the context of the litigation involving Uber and Ola in the Netherlands, where drivers are suing the ride-hailing companies to force them to release, in response to requests made under data protection law, more granular data that would allow them to coordinate for collective bargaining[37].

To avoid terminological confusion, it should be noted that these cases do not involve union using workers' data as a *bargaining chip* in the negotiation: specifically, in the case of Youtube pressure was made primarily through letters and the announcement of a "collective warning" strike (i.e., a threat to refrain from publishing content for several weeks). Nevertheless, in both scenarios it is clear that data is essential for workers to exchange knowledge and gain collective awareness of the problems they face individually, which ultimately allows them to make more substantiated requests for concessions. Thus, while on the one hand it is evident that access to data is a battling ground for labor unions, it remains to be seen whether users of a particular platform or software system can be effective in withholding the supply of data to the platform in exchange for improved conditions. The recent saga involving WhatsApp and the rolling out of its new privacy policy seemingly provides evidence that these movements could work. In this case, the leading consumer communication application announced on 4 January 2021 a take-it-or-leave-it update to its privacy policy, which involved the sharing of some information with its mother company Facebook[38]. This caused a global outcry and migration of users to competing applications, which ultimately led to a retreat by WhatsApp on the date by which users would be required to accept the new changes in order to continue using WhatsApp's services[39]. If WhatsApp's postponement of this deadline is a concession to meet the concerns

[34] Payal Arora, Linnea Holter Thompson, Crowdsourcing as a Platform for Digital Labor Unions, 12 International Journal of Communication (2018)
[35] Annemarie Kern and Valentin Niebler "Organising YouTube. A novel case of platform worker organising", Friedrich Ebert-Stiftung Institute (2020)
[36] Id., at 5
[37] ADCU, "Uber and Ola Cabs in legal bid to curtail worker digital rights and supress union organised data trusts" (ADCU, 16 December 2020) <https://www.adcu.org.uk/news-posts/uber-and-ola-cabs-in-legal-bid-to-curtail-worker-digital-rights-and-supress-union-organised-data-trusts> accessed 01 June 2021
[38] Nick Statt, "WhatsApp clarifies it's not giving all your data to Facebook after surge in Signal and Telegram users" (The Verge, 12 January 2021) <https://www.theverge.com/2021/1/12/22226792/whatsapp-privacy-policy-response-signal-telegram-controversy-clarification> accessed 01 June 2021
[39] WhatsApp, "Giving More Time For Our Recent Update" (WhatsApp, 15 January 2021) <https://blog.whatsapp.com/giving-more-time-for-our-recent-update> accessed 01 June 2021

expressed by consumer groups, despite the absence of a dedicated negotiation process, this is a promising sign for data unions.

At the same time, the success of these spontaneous initiatives suggests that it is reasonable interpret the concept of "data union" more loosely: not merely as an institutionalized collective labor movement, but more generally as a coordination by individuals to improve the conditions under which they make their clickstream data (which could be considered as the output of their labor) accessible to third parties. A tool that enables that is the protocol built on the Streamr Network, which offers the ability for people to easily bundle and sell their real-time data through its Marketplace, in exchange for tokens (which can be redeemed for real monetary value)[40]. Specific applications of this protocol are Swash, in relation to browsing data[41]; MyDiem, in relation to phone device and app usage information paired with demographic data, such as age bracket, country/region, and gender[42]; and Tracey, in relation to seafood product data as it transitions through the whole supply chain[43]. The protocol even allows consumers to pool and sell consumption data generated through existing applications: for instance, Spotify's streaming data[44]. Similar applications can be developed on the Solid protocol, which allows governments and large organizations to build personal online data stores (Pods) where users can store their data collected from multiple sources and grant permissioned access to third parties[45]. Obviously, a crucial point for the success of these applications will be the extent to which they implement adequate safeguards (in particular to prevent user re-identification and misuse of data) to prevent inadequate use of personal data. Assuming this is properly accounted for, however, these applications offer a clear advantage compared to the existing *status quo*: through these applications, users are able to reduce the asymmetry in bargaining power in data transactions, as the joint value of various data points is worth significantly more to its acquirers than the sum of its constituent parts[46].

Data pools

Another type of solution for collaborative enterprises is offered by so called "data pools", which have been described by Lindqvist as arrangements where "firms agree to share their digitalised information regarding a given market, in reference to a given service or generally in an industry, or within an e-ecosystem"[47]. The same author acknowledges, however, that this is a broad concept that may be configured in different forms: in certain cases, data pools may

[40] See Tom Hamilton, "What are Data Unions? How do they work? Which ones can I use?" (Medium, 31 March 2021) <https://medium.com/streamrblog/what-are-data-unions-how-do-they-work-which-ones-can-i-use-887e67fb7716> accessed 06 June 2021

[41] See: Swash app. Home. [online]. <https://swashapp.io/> accessed 10 June 2021. It works as a data collector, paying the user that provides the information.

[42] See: Diem Association. Home. [online]. <https://www.diem.com/en-us/> accessed 10 June 2021

[43] See: TX Project. Home. [online]. <https://tx.company/projects/tracey/> accessed 10 June 2021

[44] Tom Hamilton, supra n. 40.

[45] Tim Bernes Lee, "How do we make the circular economy go round?" (New York Times, 10 January 2021) <https://www.nytimes.com/2021/01/10/technology/tim-berners-lee-privacy-internet.html> accessed 01 June 2021

[46] See in this sense Bietti, Elettra, Locked-in Data Production: User Dignity and Capture in the Platform Economy (October 14, 2019). Available at SSRN: https://ssrn.com/abstract=3469819 or http://dx.doi.org/10.2139/ssrn.3469819

[47] Bjorn Lundqvist, Data Collaboration, Pooling and Hoarding under Competition Law (2018). Faculty of Law, Stockholm University Research Paper No. 61, Available at SSRN: https://ssrn.com/abstract=3278578

simply contain user or consumer data; while others may include technology, product or distribution information based on specific proprietary pool technology[48].

What seems to be constant in these collaborations is the sharing of industry knowledge, which at the minimum includes consumption, production or distribution data at some level of generality, and on some occasions must include technology as the data may be intelligible only through proprietary software (for instance, a user-facing database that must be licensed in order to access the information or to derive insights from raw data). Industrial data pools, where manufacturers share product information[49] and often a common ontology for reference architectures[50] or faster insights correlation[51], are a classic example of this kind of arrangements. Yet it is worth noting that the World Economic Forum speaks of a broader paradigm shift, not limited to industrial applications, but with an all-encompassing effect on the economy through shared repositories of data, analytics and machine-learning algorithms which leads to new revenue streams, products and services, as well as richer insights for a broader range of stakeholders[52].

Data trusts

A fourth category that is fraught with definitional difficulties is that of "data trust", as "trust" itself as a legal characterization encompasses a range of related concepts[53]. In general terms, it is a form of property holding where the receiver (trustee) must apply the property exclusively for the benefit of someone else (beneficiary), whose identity may or may not coincide with that of the person who set up the trust (settlor).

Other than by natural and legal persons, trusts may be the result of judicial or legislative creation. In all these cases however, independently from their origin, trustees have two different duties towards their beneficiary: first, a fiduciary duty of undivided loyalty, which implies that the trustee must act in the sole interest of the beneficiary (and therefore cannot assume any conflicting interest); second, a duty of care, which implies that they should exercise reasonable care in the handling of the property.

Recent work has applied these characteristics to arrangements whereby one or more individuals are entrusted with the holding of data, arguing that this instrument is flexible enough to permit the emergence of a variety of trusts, each of which would negotiate with data collectors on the beneficiary's behalf in accordance with the rules and values enshrined in each particular governing statute[54]. In some cases, data trustees can also be public entities that are assigned with the responsibility to manage datasets (in particular in the health sector) in accordance with

[48] Id., at 7
[49] This is the case, for instance, of the voluntary partnership between enterprises promoted and facilitated by the Japanese Minister for Economy, Trade and Industry: see https://www.meti.go.jp/english/press/2017/0328_003.html
[50] See for instance the Internet Industry Consortium, at http://www.iiconsortium.org/
[51] See for instance the Open Manufacturing Platform, at https://open-manufacturing.org
[52] WEF, A New Paradigm for Business of Data, Briefing Paper (July 2020)
[53] Charlie Webb & Tim Akkouh, *Trusts Law* (Fifth ed., Palgrave 2017), p. 1
[54] Sylvie Delacroix, Neil D Lawrence, Bottom-up data Trusts: disturbing the 'one size fits all' approach to data governance, International Data Privacy Law, Volume 9, Issue 4, November 2019, Pages 236–252, https://doi.org/10.1093/idpl/ipz014

predefined criteria, often including granting vetted access to third parties, and therefore play a role that is also known in some jurisdictions as one of "data custodian"[55].

Regardless of the formal qualification, the important feature from the standpoint of competitive analysis is that they lead to potentially powerful entities managing the interests of various groups of society. This can have positive effects in allowing their beneficiaries to overcome a weak bargaining position and be heard more effectively within society, but also the negative effect of potentially extending the effects of market power held by any of the represented stakeholder groups through a delegated form of control over the undertaking. More case-by-case discussion is needed, as pointed in section 3 below.

Data exchanges

The simplest form of collaboration that we can identify is within the context of a so-called "data exchange". In broad terms, exchanges are intermediary platforms that facilitate transactions between data suppliers (or data holders) and data consumers (or data users) with different levels of interaction between parties: either on a one-to-one basis, or on a "club" basis, or else on an "open to all" basis.[56] One example that enables all the three forms of exchange is Dawex, which has its own open marketplace (the largest in the world) and also offers its own infrastructure to allow companies to set up a "data exchange platform" where they can distribute, source, exchange, share and commercialize data and/or orchestrate data ecosystem.[57]

These exchanges may involve personal data, therefore triggering the application of EU data protection rules, but we assume for simplicity that data protection, privacy and cybersecurity concerns are duly taken care of. However, different considerations altogether apply when the focus of the transaction is personal data about a specific individual, who is the data holder in the transaction: in such case we speak about a "Personal Data Exchange"[58], i.e. a technology platform that enables individuals to transfer their personal data to a data user in exchange for some benefit. These benefits can include insights, a more personalized service, a discount or even monetary value. Examples are offered by pioneering "personal management systems" or

[55] See, for instance, Western Australia´s Data Stewardship and Custodianship Policy (2016), available at <https://ww2.health.wa.gov.au/~/media/Files/Corporate/Policy%20Frameworks/Information%20management/Policy/Data%20Stewardship%20and%20Custodianship%20Policy/Data-Stewardship-and-Custodianship-Policy.pdf> accessed 1 December 2020. The definition for data custodian refers to: " The person(s) responsible for the day-to-day management of a data collection, as nominated by the Data Steward. Data Custodians assist the Data Steward to protect the privacy, security and confidentiality of information within data collections. Data Custodians also aim to improve the accuracy, usability and accessibility of data within the data collection". In turn, the definition provided for "Data Steward" is :" A position with delegated responsibility from the Director General of the Department to manage a data collection. The Data Steward's primary responsibility is to protect the privacy, security and confidentiality of information within data collections. Data Stewards also approve the conditions for appropriate use and disclosure of information for clearly defined purposes that comply with WA Health's statutory obligations and Information Management Policy Framework."

[56] Oliver Wagner, The Rise of Data Exchanges Frictionless Integration of Third-Party Data, Eckerson Group Report White Paper (September 2020). Available at <https://www.harbrdata.com/wp-content/uploads/2020/10/101220_Data_Exchanges_Harbr.pdf> accessed 1 June 2021.

[57] See: Dawex Systems. <https://www.dawex.com/en/data-exchange-platform/>

[58] For an early definition, see Michael Haupt, Introducing Personal Data Exchanges & the Personal Data Economy, Medium (7 December 2016), available at <https://medium.com/project-2030/what-is-a-personal-data-exchange-256bcd5bf447> accessed 1 June 2021

"personal data stores" like Digi.me[59], Citizenme[60], Meeco[61], DataCoup[62], and several others that have propped up in recent years. It should also be mentioned that personal data exchanges do not necessarily involve the transfer of personal data: in some instances, individuals can simply use the exchanges to be matched with marketers or researchers who have an interest in understanding a particular demographic, without the need or possibility to identify individuals. UBDI is an example of a platform specialized in such exchanges, and which combines the exchanges with a system of verification of attributes that can be used to receive offers to participate in tailored studies[63].

For these kinds of scenarios, it is said that companies on the exchange are trading in insights, rather than data[64]; but the line between these two is a fine one, as detailed insights could be used (together with publicly available data) to identify individuals. The system of exchange used for behavioral advertising is a notable example of this, which has led to complaints lodged with the data protection authorities of Ireland, United Kingdom and Poland[65]. This is just one example highlighting the importance for data exchanges to implement key principles of data minimization and data privacy by design, in order to prevent undesirable downstream effects on individuals that may otherwise be identified.

2. The EU vision of "European data spaces"

In its recent Communication on an EU Data Strategy, the Commission lays out its vision to "enable the EU to obtain by 2030 a share of the the data economy – data stored, processed and put to valuable use in Europe – that at least corresponds to its economic weight"[66]. This vision is built around the idea of a single European data space, also defined as a "genuine single market for data" where personal and non-personal data is securely stored and made "easily available" to third parties, while in full respect of data protection and other relevant legislation. Specifically, the European data space will be composed of a secure IT environment for processing of data by an open number of organisations, and a set of rules of legislative, administrative and contractual nature that determine the rights of access to and processing of the data, which allows data to be made available on a voluntary basis and reused against remuneration or for free, depending on the data holder's decision[67].

It is worth stressing the particular emphasis of this initiative, following in the footprint of successful experiences with data spaces at the national level[68], is one of creating a framework that not only facilitates data use in compliance with existing legislation, but also ensures effective enforcement of EU law. The latter, presumably a reference to the location of the

[59] See: Digi.me LTD. <https://digi.me/> accessed 1 June 2021
[60] See: Citzenme. <https://www.citizenme.com/> accessed 1 June 2021
[61] See: MEECO Inc. <http://meeco.com/meeco/> accessed 1 June 2021
[62] See: DataCoup Inc. <https://datacoup.com/> accessed 1 June 2021
[63]See: UBDI. Home [online] <https://www.ubdi.com/> accessed 1 June 2021
[64] Taylor, Linnet, et al. "(re)making Data Markets: An Exploration of the Regulatory Challenges." SocArXiv, 29 Oct. 2020. Web. Available at <https://osf.io/preprints/socarxiv/pv98s/> accessed 1 June 2021, p. 12
[65] Katarzyna Szymielewicz, "Legal battle over online behavioural advertising widening" (Policy Review, 30 January 2019) <https://policyreview.info/articles/news/legal-battle-over-online-behavioural-advertising-widening/1384> accessed 06 June 2021
[66] European Data Strategy, supra n. 6
[67] Impact Assessment of the DGA, p. 8.
[68] Examples are Finnish Health and Social Data Permit Authority (https://www.findata.fi/en/), the French Health Data Hub (https://www.health-data-hub.fr/), and the German Forschungsdatenzentrum (https://www.forschungsdatenzentrum.de/en).

infrastructure for such data spaces[69], is not elaborated further in the Communication. By contrast, the Commission does go into some level of detail regarding the issues that it aims to address in order to stimulate such data use: these include prioritizing interoperability requirements and standards within and across sectors; strengthening governance mechanisms by involving public and private actors; facilitating decisions over the use of data for research purposes; and making it easier for individuals to allow the use of data for public good (so called "data altruism").

It is not surprising that, given the potential tensions between cooperation and competition, the Communication specifically mentions competition law. In particular, the Commission promises to give more guidance to stakeholders on the compliance of data sharing and pooling arrangements with EU competition law by means of an update of the Horizontal Cooperation Guidelines[70], in addition to being prepared to provide support decisions for specific situations on what data can be used and under what standards[71]. This certainly demonstrates an important commitment to the data-driven economy; at the same time, the absence of a specific implementation strategy leaves many companies in the dark regarding the legal treatment of data collaboratives. Before sketching ways in which existing competition law would see these initiatives, it is useful to break down the 9 types of data spaces that the Commission envisages in its Communication on a European Strategy for Data:

- A Common European industrial (manufacturing) data space, to support the competitiveness and performance of the EU's industry, by gathering key players from the manufacturing sector and have them agree about the conditions under which they would be ready to share their data and how to further boost data generation, notably via smart connected products.
- A Common European Green Deal data space, to use the major potential of data in support of the Green Deal priority actions on climate change, circular economy, zero-pollution, biodiversity, deforestation and compliance assurance. In particular, the Commission will launch a "GreenData4All" initiative to assist in collecting, sharing, processing and analysing large volumes of data relevant for assuring compliance with environmental legislation and rules related to the priority actions set in the Green Deal; and "Destination Earth", a digital modelling platform to visualize, monitor and forecast natural and human activity on the planet in support of sustainable development.
- A Common European mobility data space, to position Europe at the forefront of the development of an intelligent transport system, including connected cars as well as airline, rail and navigation. Such data space will facilitate access, pooling and sharing of data from existing and future transport and mobility databases.
- A Common European health data space, to strengthen and extend the use and re-use of health data, including helping healthcare authorities to take evidence-based decisions to improve the accessibility, effectiveness and sustainability of the healthcare systems; and deploy the data infrastructures, tools and computing capacity for the European

[69] This was seemingly confirmed by the localization requirement that was originally attached to the providers of "data sharing services" pursuant to a leaked version of the Digital Governance Act. However, this requirement disappeared from the text of the Act that was published on 2 December 2020. See: MyData, <https://mydata.org/wp-content/uploads/sites/5/2020/11/datagovernanceact_oct28_leak.pdf>

[70] Communication from the Commission — Guidelines on the applicability of Article 101 of the Treaty on the Functioning of the European Union to horizontal co-operation agreements Text with EEA relevance
OJ C 11, 14.1.2011, p. 1–72, at 14

[71] *Id.*

health data space, more specifically support the development of national electronic health records (EHRs).

- A Common European financial data space, to stimulate, through enhanced data sharing, innovation, market transparency, sustainable finance, as well as access to finance for European businesses and a more integrated market.
- A Common European energy data space, to promote a stronger availability and cross-sector sharing of data, in a customer-centric, interoperable, secure and trustworthy manner, as this would facilitate innovative solutions and support the decarbonisation of the energy system.
- A Common European agriculture data space, to enhance the sustainability performance and competitiveness of the agricultural sector through the processing and analysis of production and other data, allowing for precise and tailored application of production approaches at farm level; and support the emergence of an innovative data-driven ecosystem based on fair contractual relations as well as strengthen the capacities for monitoring and implementing common policies and reducing administrative burden for
- Common European data spaces for public administration, in order to improve transparency and accountability of public spending and spending quality, fighting corruption, both at EU and national level, and to address law enforcement needs and support the effective application of EU law and enable innovative 'gov tech', 'reg tech' and 'legal tech' applications.
- A Common European skills data space, to reduce the skills mismatches between the education and training system on the one hand and the labour market needs on the other, including by facilitating secure and interoperable credentials in digital format.
- A European Open Science Cloud, which provides seamless access and reliable re-use of research data to European researchers, innovators, companies and citizens through a trusted and open distributed data environment and related service

As this list attests, the European Commission´s ambition is to give rise to a wide range of collaborations, often involving both private and public actors, by facilitating the sharing of data already produced and held by existing players. The Commission sees the role of public sector as pivotal in this respect, as it already counts on a dedicated framework facilitating the re-use of data held by public sector bodies and public undertakings[72]. Perhaps most notably, the EU legislator has already introduced the obligation for these entities to make re-use available for free, in machine-readable format, via APIs, and where relevant with bulk download, when it comes to so-called "high-value datasets"[73]. The PSI Directive called the Commission to identify such datasets with a delegated act within the following six categories: (1) geospatial; (2) earth observation and environment; (3) meteorological; (4) statistics; (5) companies and company ownership; and (6) mobility. It also determined the criteria that must be followed to identify such databases, which are built around their potential to (a) generate significant socioeconomic or environmental benefits and innovative services; (b) benefit a high number of users, in particular SMEs; (c) assist in generating revenues; and (d) be combined with other datasets[74].

[72] See below, section 3.2

[73] Directive (EU) 2019/1024 of the European Parliament and of the Council on open data and the re-use of public sector information (2019), Art. 14

[74] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and

Yet, the Communication does not set out what legal or organizational form will be chosen for the implementation of EU data spaces. For the most part, data spaces appear to be dependent on producers´ cooperation, which could be supported by virtually all the models described in section 2, with the possible exception of data unions. However, it is reasonable to expect that the adoption of one model over the other will depend on the specific needs of each of this sectors or areas of focus. For example, it is important to distinguish initiatives that involve the processing of personal data, and therefore will need to be conducted in compliance with the GDPR and any relevant national legislation. This may be relevant for the European Mobility Data Space, the European Health Data Space, the European Energy Data Space, and the European Open Science Cloud: in all these cases, the data collaboratives should take into account the need for a legal basis and appropriate safeguards that justify the processing of information relating to an identified or identifiable natural person ('data subject')[75]. In the other data spaces, the risk of identification is lower, though it cannot be entirely ruled out. As a result, an important responsibility of undertakings joining the data space will be to assess and possibly certify compliance of their services with the requirements of the applicable data protection regulation.

A further observation is that the success of these initiatives depends to a large degree on voluntary actions by market participants, making it particularly crucial to understand what would trigger this voluntary behavior. This is addressed in passing by the Communication: the European Commission´s vision is that the organisations contributing data would get a return in the form of increased access to data of other contributors, analytical results from the data pool, services such as predictive maintenance services, or licence fees[76]. While the breadth of this exchange does not prejudge the types of business models that firms may adopt within data spaces, this vision begs at least two questions about plausible future: first, will such an incentive structure be sufficient for players holding key datasets to make them available to competitors and other third parties? Second, and alternatively, is this an effective strategy to encourage collaboration of smaller industry players, and thereby create a level playing field that allows new entrants to challenge more powerful and data-reach incumbents? On both issues, the European Commission seems to put great faith in the potential of openness and collaboration, discounting the risks of anticompetitive behaviour by assuring the compliant nature of these initiatives. In that respect, many may be left wondering how this compliance can be assured, given the lack of details provided criteria in the applicable legal framework. The following section sketches some of the key interpretative issues that might arise in the context of data spaces.

3. **Examining collaborations under the existing framework**

Although EU Competition law has a strict policy against cartels (also known as "hard-core" horizontal agreements), it does not necessarily view competitor collaboration with disfavour. Indeed, it encourages it through specific guidelines and block exemption regulations (see below, sections 3.1 and 3.2). However, collaboration is seen with suspicion because it is a vehicle of collusion: due to the secretive nature of cartels and the difficulty of finding hard evidence of agreements, the scope of application of art 101 TFEU includes conduct that may be considered to lie at the periphery of a price-fixing or market allocation scheme. This is done through the notion of "concerted practice", famously defined from the *Dyestuff* case in 1992 as

repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance) OJ L 119, 4.5.2016, p. 1–88, Art. 14

[75] *Id.,* Art. 4

[76] European Data Strategy, p. 6

"a form of coordination between undertakings which, without having reached the stage where an agreement properly so-called has been concluded, knowingly substitute practical cooperation between them for the risks of competition"[77]. This concept is sufficiently broad to cover a very wide range of activities, as long as they involve "*any* direct or indirect *contact* between operators by which an undertaking *may influence the conduct* on the market of its actual or potential competitors *or disclose to them its decisions or intentions* concerning its own conduct on the market[78]. The breadth of this definition suggests that collaborators should thread carefully with the type of conduct they require or encourage from each other, in particular for conduct that is liable to produce effects equivalent to price-fixing or market-sharing.

In the context of data collaboratives, it is not difficult to imagine scenarios where such effects are involved: for instance, this would be the case when a data pool or a data exchange determines, or even merely recommends, the fees to be charged by data holders for granting access to certain kind of data; or when one or more data cooperative coordinate their activities establishing restrictions to membership according to locality. Whether such effects are actually produced in the market is irrelevant[79], as these practices are condemned as restrictions of competition "by object". The presumption of anti-competitiveness intrinsic in this characterization can be rebutted by providing an objective justification for what *prima facie* falls foul of the provision[80]: for instance, the necessity of price-fixing to enable the emergence of a product which would otherwise not be produced, or the need to divide groups of engagement in order to attend to local regulations. However, it is important to highlight that objective justifications cannot be invoked simply to privately enforce compliance with legal requirements: the European Court of Justice has clarified that this is a task for public authorities, not private undertakings[81].

In addition to direct participation to collusive schemes between competitors, article 101 TFEU applies to third parties who facilitate such schemes. For instance, in *AC-Treuhand*, the defendant was found liable, despite its lack of participation in the cartel, due to fact that it had been entrusted as an agent to a number of organizational and administrative tasks that were instrumental to the formation and coordination of the cartel, of which it was aware[82]. The Court of Justice also clarified in *Heat Stabilisers* that the requisite intent for the facilitator does not raise a high bar, as it is established not only when the facilitator was aware of the conduct planned or put into effect by the other undertakings, but also when it could reasonably have foreseen it and was prepared to take the risk[83]. Transposing this to the data space context, it is not difficult to imagine that the technological and organizational structure used in fulfilment of

[77] *ICI v Commission*, Case 46/89 EU: C: 1972:70, para. 64.
[78] *T-Mobile,* 2009 E.C.R. I-04529, para 33 (emphasis added).
[79] *Id.,* paras 51-53
[80] For example, the need for collective action to fight copyright infringement was accepted as a justification for the coordination of collecting societies. See Case T-442/08, e al., *CISAC v Commission*, EU:T:2013:188, paras 87 ff.
[81] Case C-68/12, *Protimonopolný úrad Slovenskej republiky v Slovenská sporiteľňa a.s*., EU:C:2013:71, para. 20
[82] "Inter alia, storing certain secret documents relating to the cartel […] on their premises; collecting and treating certain information concerning the commercial activity of the three organic peroxide producers; communicating to them the data thus treated; and completing logistical and clerical-administrative tasks associated with the organisation of meetings between those producers […], such as the reservation of rooms and the reimbursement of their representatives' travel costs". Case T-99/04, *AC-Treuhand AG v Commission,* EU:T:2008:256, para 2.
[83] Case C- 194/14 P, *AG Treuhand v Commission*, EU:C:2015:717, para 30.

the needs of a data space may raise competitive concerns, particularly to the extent that it does not incorporate appropriate safeguards to prevent collusion among its members.

As this overview shows, both participation to and coordination of data spaces can involve activities that trigger the application of competition law, thus potentially turning into a mined field to the untrained eye. European institutions have tried to address some of the complexity involved the application of article 101 TFEU by adopting legal instruments that give more detailed guidance to stakeholders, but there are questions as to the suitability of these instruments to capture the specificities of data collaboratives.

3.1. Guidelines on horizontal cooperation agreements

A first comprehensive set of guidelines can be found the Guidelines on horizontal cooperation agreements from 2011[84], replacing the previous version from 2001[85]. These apply not only to initiatives involving actual and potential competitors, but virtually any kind of collaborations between firms operating at the same level of the production chain, regardless of the existence of competitive pressure (for example, firms that offer similar products but for which there is no substitability[86]). The goal of the Guidelines is to provide an analytical framework for the most common types of horizontal co-operation agreements, sketching the important elements for a competition law analysis of the following: research and development agreements; production agreements, including subcontracting and specialisation agreements; purchasing agreements; commercialisation agreements; standardisation agreements including standard contracts; and information exchange[87]. Given the amount of potentially overlapping practices covered by this guidance, it should not surprise that the first problem that is addressed early on in the text is a jurisdictional one: specifically, what rules should be used as a reference in the presence of complex operations that involve several types of agreements?

The criteria provided by the Guidelines to answer these questions, by determining the so-called "centre of gravity", refer to two important factors: the starting point of cooperation and the level of integration of the different functions which are combined. In practice, the concept of "starting point" means that parties should identify the core element of cooperation (for instance, the conduct of joint research and development) and whether any further cooperation (for instance, joint selling) is subordinate to that, in which case any subordinate activities will be subsumed in the analysis of the core elements. However, the concept of "integration" of functions can shift the centre of gravity to another category (for instance, production agreements) to the extent that it leads to joint ownership and control on the results of the cooperation; and note that this is irrespective of whether it is undertaken in conjunction with some other form of cooperation (for instance, R&D)[88]. For this reason, it is particularly important to understand the legal and organizational structure chosen by a data collaborative: the closer these initiatives are to the concept of a cooperative, the more holistic the analysis

[84] Communication from the Commission — Guidelines on the applicability of Article 101 of the Treaty on the Functioning of the European Union to horizontal co-operation agreements Text with EEA relevance, OJ C 11, 14.1.2011, p. 1–72
[85] Commission Notice — Guidelines on the applicability of Article 81 of the EC Treaty to horizontal cooperation agreements, OJ C 3, 6.1.2001, p. 2–30
[86] See para. 1 of the Guidelines, that refer to "Non-competitors" and offer the example of companies operating in the same product market but a different geographical area.
[87] Para. 5
[88] Para. 14

and the less necessary it will be to examine its constituent parts under different categories of agreements.

As a guiding principle, the Guidelines recognize the enormous benefits that can come from collaboration, in particular if they combine complementary activities, skills or assets[89].At the same time, they identify possible concerns in the loss of competitive pressure between the parties to the cooperation and *vis a vis* third parties, as well as the collusion that can result from the disclosure of sensitive information or the achievement of substantial commonality of costs[90]. Whether this will actually happen depends on a number of factors, including the market power of the involved parties, the nature and content of the agreement, and the extent to which it contributes to the creation, maintenance or strengthening of that market power or allows the parties to exploit such market power. In this regard, it is important to stress that although the Guidelines refer to market shares when establishing safe harbors, they contain an introductory warning that the use of this indicator may not be appropriate in certain cases, particularly in light of the level of concentration and the position of the parties in the market[91]. Indeed, for the assessment of market power of the collaborating forms it may be necessary to take into account the stability of market shares over time, the existence of entry barriers and the likelihood of market entry, as well as any countervailing power of buyer[92]. This is particularly relevant in the context of data collaboratives, as the concept of data accumulation is not neatly fitting into traditional indicators and methods for the estimation of market power. While Gal and Rubinfield[93] argue that before establishing market power it may be necessary to consider the existence of barriers to entry at other levels of the data value chain, such as data storage and analysis, Graef suggested that this is more likely in the presence of four conditions: (1) data is a significant input into the service delivered by the online platform; (2) the incumbent relies upon IP law to deny access to the relevant dataset; (3) there are no or few substitutes available to potential new entrants; and (4) it is unviable for competitors to self-collect data to build a competing dataset[94]. To this, one can add that the importance of data for competitiveness in a given market can also undermine the countervailing force of multi-homing, at least in the absence of sufficient volumes, and should therefore not be underestimated. Summing up, it is clear that no silver bullet can be used to estimate the importance of a particular dataset in a given market: as acknowledged by the Franco-German Report on Competition and Data and several other commentators[95], it is a context-specific question, to be answered on a case-by-case basis.

Information exchange

The first category examined in the Guidelines is a general one, which can form part of a range of possible cooperation agreements: information exchange. The reason why this common practice deserves particular attention is that it may be considered sufficient to qualify as a

[89] Para. 2
[90] Para. 33-35
[91] Para. 45
[92] Para. 46
[93] Michal S. Gal and Daniel L. Rubinfeld, "Access Barriers to Big Data" (Arizona Law Review, Vol. 59:339, 2017), pp. 341-381.
[94] Graef, Inge, EU Competition Law, Data Protection and Online Platforms: Data as Essential Facility (PhD summary) (June 29, 2016). EU Competition Law, Data Protection and Online Platforms: Data as Essential Facility, Kluwer Law International 2016, Available at SSRN: https://ssrn.com/abstract=3635378
[95] E.g. Gal and Rubinfield, supra n. 93; D. Daniel Sokol, "Does Antitrust Have a Role to Play in Regulating Big Data?", in R. Blair & D. Sokol, The Cambridge Handbook of Antitrust, Intellectual Property, and High Tech, pp. 293-316

concerted practice, and thus a restriction of competition, even "by object". This is the case for disclosure of individualized future prices, and EU competition law applies a presumption to such cases because exchange of information removes "strategic uncertainty" about future conduct in the market[96]. In this scenario, the addressees of the message are deemed to have participated in the collusive scheme as long as they were aware or ought to have been aware of such message and did not publicly take distance from it, or act contrary to the recommended practice in the market[97].

This means that information exchanged in a data space may give rise to liability for all participants if it reveals strategic details about their future conduct on the market. For instance, at least in principle, information concerning a dataset in the open manufacturing space could permit beneficiaries to estimate the amount of raw material that a particular manufacturer needs to buy at regular intervals, and their processing time. Is this sufficiently strategic? It appears so: by way of example, the Guidelines cite prices and quantities, customer lists, production costs, quantities, turnovers, sales, capacities, qualities, marketing plans, risks, investments, technologies and R&D programs and their results[98]. However, ultimately the strategic usefulness of such data will depend on the specific way in which undertakings compete, and thus on the characteristics both of the information and of the market in which such exchange occurs. As a general guideline, the more the information is fresh, strategic (in the sense of removing strategic uncertainty), frequent, future-oriented, in individualized form and covering the entirety of the market, and the more the market is transparent, concentrated, non-complex, stable and symmetric, the more likely it is that the exchange gives rise to collusive concerns. Furthermore, in a data-driven environment, the concern might be bigger to the extent that the shared information involves specific characteristics of consumer profiles that are used to target products or services: building offers on the basis of common profiles may lead to a loss of competition, in particular as it makes it more difficult for consumers to benefit from more innovative (and perhaps more accurate) profiling techniques.

Aside from collusion, exclusionary concerns may arise when parties exchange on an exclusive basis commercially sensitive information, which enables them to foreclose competition either in the same market where these undertakings operate, or in a related market[99]. A good example of this situation is the alliance between advertisers for the purpose of targeting, whereby companies share with advertising platforms like Google and Facebook the information they have about their customers in order to better target them on those platforms[100]. The characteristics of data-driven innovation are such that the consequences of greater intelligence (over customers and competitors) can provide an advantage in markets that are distant from the one in relation to which that information was provided: for instance, Facebook could use information about the ad campaigns made by your local supermarket to advertise specific types of food delivery options near you[101].

However, one should not haste to the conclusion that exchange of sensitive or strategic information is necessarily anticompetitive. In fact, significant efficiencies can be generated by

[96] See *T- Mobile*, supra n.78
[97] "*Eturas" UAB and Others v Lietuvos Respublikos konkurencijos taryba*, Case C-74/14 (2014), para. 49
[98] Horizontal Guidelines, supra n. 70 , para. 86
[99] Id., para. 70-71
[100] ACCC, Digital Advertising, Interim Report, p. 60
[101] This seems to be the theory advanced by the European Commission in its investigation against Facebook on 4 June 2021, where it argued that the company used data from advertising to gain an advantage into its marketplace. See https://ec.europa.eu/commission/presscorner/detail/en/ip_21_2848

an exchange of information, particularly when it comes to consumer data exchanged between companies in markets with asymmetric information about consumers. The leading example in this sense is *Asnef Equifax*[102], where the ECJ ruled, on a preliminary reference from an Austrian court, in favor of the legality of a register where banks pooled creditworthiness information about their customers. The Court noted that a register system such as the one under examination can improve the functioning of the credit market, including enabling the entry of new competitors and improving the mobility of consumers[103]. It went on to clarify that an assessment of legality can only be made after looking at the context of the agreement, which in that case included three important considerations: first, that the market was fragmented (not concentrated), thus making it more difficult to learn on that basis the market position and commercial strategy of competitors[104] ; second, that the information about the identity of the lenders was not revealed, thus reinforcing the difficulty of tracing information about lending to a specific competitor[105]; and third, the information in the register was available to all the operators in the market in a non-discriminatory manner, making the market more contestable[106].

Notwithstanding the fact that the Court in *Asnef Equifax* found the register not to restrict competition, it also answered the question posed to it by the Austrian court regarding the applicability of article 101 (3). The crucial point in this regard was the second element of the article 101 (3) test, i.e. the existence of a "fair share of the resulting benefits" for consumers, in particular because on the basis of the register some consumers might only receive credit at worse conditions or be refused credit altogether. Nevertheless, the Court simply reminded that it is the interest of all consumers in the relevant market that should be taken into account, not that of a particular subset of it[107]: in other words, welfare trade-offs between consumers of the same market are admitted. Unfortunately, the national court did not provide a quantification of the share of consumers that would be adversely impacted, which would arguably have triggered a more detailed and convincing reasoning by the Court of Justice. To what extent should competition law prevent cream-skimming? The answer to this question is crucial in markets that are not covered by universal service obligations, and would provide important guidance for situations where profiling enables discrimination not merely of price, but also, of products and services offered to consumers. It has been argued, for instance, that personalized exploitation should be recognized as consumer harm in antitrust analysis[108], which would be supported by the Commission´s practice to define separate markets where a distinct group of consumers could be subject to price discrimination[109].

R&D Agreements

A second category addressed in the Guidelines is research and development agreements, which concern a joint effort to research and develop new products or technologies, and possibly a joint exploitation of the results of such collaboration. Here, it should be borne in mind that a specific exemption to the application of article 101 TFEU is provided by the Block Exemption

[102] Case C-238/05 *Asnef/Equifax* ECLI:EU:C:2006:734
[103] *Id.,* para. 56-57
[104] *Id.,* para. 58
[105] *Id.,* para. 59
[106] *Id.,* para. 60
[107] *Id.,* para. 70
[108] Inge Graef, "Consumer sovereignty and competition law: From personalization to diversity" (Common Market Law Review, Volume 58, Issue 2, 2021) pp. 471-504
[109] Commission Notice on the definition of relevant market for the purposes of Community competition law, O.J. 1997, C 372/5, para 43

Regulation on R&D, which creates a carve-out for collaborations among competitors whose total market share does not exceed 25%, or among non-competitors, as long as the parties to the agreement have full access to the final results and there are no "hard core" restrictions[110].

This safe harbor creates a zone of enhanced legal certainty for the self-assessment of data cooperation. However, its relevance for data collaboratives should not be overstated: the exemption only applies to provisions which relate to the assignment or licensing of intellectual property rights to one or more of the parties or to an entity the parties establish to carry out the joint research and development, paid-for research and development or joint exploitation; and only provided that those provisions do not constitute the primary object of such agreements (as otherwise this would be dealt with under the regulation on technology transfer[111]), but are directly related to and necessary for their implementation.[112] In this regard, it should be noted that, while intellectual property can be an important component and driver of data collaborations, it need not be: despite the breadth of IP rights in data (especially, database rights)[113], collaborating firms might not have claimed such rights at the start of their partnership. They may have structured their agreements without linking to intellectual property the value that each participant brings, rather referring to granting access to digital assets in exchange for some type of benefit in the final results. In this case, it is clear that collaborations take place outside this safe harbor, and as a result, they will be governed by a different instrument: the Guidelines. The same conclusion can be reached with regard to any clauses that the Block Exemption Regulation defines as "excluded restrictions", one of which concerns the obligation not to challenge the validity of any intellectual property held by the parties that is relevant to the R&D[114].

The Guidelines go a long way in clarifying the analytical steps to be observed for the assessment of R&D, despite the challenge of the moving boundaries between existing and future markets. However, the main complexity in their application to data collaborations lies in the anticipation of the possible effects of such agreements: not just for existing product or

[110] See Commission Regulation (EU) No 1217/2010 of 14 December 2010 on the application of Article 101(3) of the Treaty on the Functioning of the European Union to certain categories of research and development agreements, Art 5.Those hardcore restrictions concern the following: clauses that prevent engagement in R&D in unrelated fields, or in the same or related field after completion of the agreement; sales or output limitations, with the exception of sales and production targets, practices constituting specialisation in the context of exploitation, and the restriction of the freedom of the parties to manufacture, sell, assign or license products, technologies or processes which compete with the contract products or contract technologies during the period for which the parties have agreed to jointly exploit the results; price fixing; restriction of passive or active sales, with the exception of the possible requirement of exclusive license to another party; requirements to refuse to meet demand from customers in the parties' respective territories, or from customers otherwise allocated between the parties; and requirement to make it difficult for users or resellers to obtain the contract products from other resellers within the internal market.

[111] Commission Regulation (EU) No 316/2014 of 21 March 2014 Ensuring technology transfer agreements respect competition rules

[112] Block Exemption Regulation on R&D, supra n. 110, Art. 2.2

[113] See Zingales, Nicolo, Of Coffee Pods, Videogames, and Missed Interoperability: Reflections for EU Governance of the Internet of Things (December 1, 2015). TILEC Discussion Paper No. 2015-026, Available at SSRN: https://ssrn.com/abstract=2707570 or http://dx.doi.org/10.2139/ssrn.2707570

[114] The two non-excluded restrictions are the following: the obligation not to challenge after completion of the research and development the validity of intellectual property rights relevant to the research and development (without prejudice to the possibility to terminate the agreement in such an event); and the obligation not to grant licenses to third parties to manufacture the contract products or to apply the contract technologies unless the agreement provides for the exploitation of the results of the joint research and development or paid-for research and development by at least one of the parties, and such exploitation takes place in the internal market vis-à-vis third parties. See Horizontal Cooperation Guidelines, supra n. 70 art. 6.

technology markets impacted by the agreement**,** but also for new markets and innovation more broadly. This assessment is more viable when the process of innovation is structured in a way that competing research poles cannot be identified at an early stage, in which case the credibility of alternative poles will be considered on the basis of the nature, scope and size of R&D efforts, their access to financial and human resources, know-how/patents or other specialized assets and the timing and capability to exploit possible results[115]. That assessment becomes more challenging, however, where competing research poles cannot be identified, in which case the Commission has committed to confine itself to the analysis of *existing* markets "absent exceptional circumstances"[116]. Whether specific processes of data-related innovation can give rise to such circumstances, or even constitute one of competing R&D poles, remains to be seen. Arguably, this is more likely to happen for R&D cooperation involving enhanced protection of personal data (something I have referred to as "data protection innovation"), rather than the shared utilization of data as a resource (which I call "data-driven innovation")[117], as in the latter case the innovation process is inherently less predictable[118].

The Guidelines recognize that restrictive effects will normally not occur if parties bring together complementary skills, technology or resources, rather than being potential competitors[119]. While this is relatively easy to determine with regard to patents, however, it might be difficult to apply in the context of pooled data: the weight of data inputs in the development of new products and services may be so significant that it turns into competitor firms that were previously operating in unconnected markets. For instance, access to a shared list of customers of a casino or an upscale restaurant might offer targeting opportunities to firms operating at different level of the value chain, which may be operating in different relevant market, but still competing for consumers' limited attention. Even if we were to reject an expansive approach to potential competition, how do we determine complementarity in this context? Datasets can be complementary for some players, and yet substitutable for others: think, for instance, about the limited significance of aggregate car usage data for an insurance provider, but its relative importance for a car repair shop.

Another important distinction in the guidelines is between pure R& D agreements and those that involve collaboration on secondary activities, for instance licensing, production and marketing: it is recognized that the latter type of agreements is more likely to give rise to anticompetitive effects, presumably due to the risk of leveraging that results in anticompetitive foreclosure. Ultimately, the assessment will depend on the market power of the concerned parties and the impact of the agreement on existing as well as on new product markets (particularly if only a limited set of R&D poles for the creation of those new products is available). Once again, this brings us back to the problem of our limited understanding of the role of data in the assessment of market power, and to our poor vocabulary in recognizing its key role as source of power in new markets: the concept of "R&D poles" strikes as ill-suited for the assessment of possible negative impacts from data accumulation on future markets, simply because such accumulation does not necessarily entail a commitment to entry a particular line of business: the non-rival nature of data is such that, from a technological

[115] *Id.*, 7
[116] *Id.,* 122
[117] Zingales, Nicolo, Data Protection Considerations in EU Competition Law: Funnel or Straightjacket for Innovation? (June 30, 2016). P. Nihoul and P. Van Cleynenbreugel, The Role of Innovation in Competition Analysis (Edward Elgar, 2018 Forthcoming), Available at SSRN: https://ssrn.com/abstract=3158008 or http://dx.doi.org/10.2139/ssrn.3158008, p. 5
[118] Victor Mayer-Schoenberger and Yann Padova, 'Regime Change? Enabling Big Data through Europe's New Data Protection Regulation' (2016)17 Columbia Science and Technology Law Review 315, 319
[119] Horizontal Cooperation Guidelines, supra n. 70, para. 130

standpoint, they can be used as inputs across a variety of product and services. A missing question in this assessment, however, pertains to the likelihood of the parties being able to do so from a legal standpoint, in particular considering the constraints imposed by data protection law. We will expand on this in section 4 below.

Production agreements

A third important category addressed by the Guidelines is production agreements, which can give rise to production of goods either in the form of a joint venture or through looser forms of cooperation involving the entrustment of another party for the production of a good (the subcontractor)[120]. Thus, within the realm of data collaborations, this type of analysis is not limited to producer cooperatives: it includes data pools and certain types of data trusts. As an example, we may think of a scenario in which data is pooled from a variety of industry partners and the responsibility for the provision of insights is assigned either to the group, or to a third party which is invited to run algorithms on the joint dataset.

With regard to this category, EU legislation provides a safe harbor for all agreements entered into between parties that do not have more than 20% of combined market share in any relevant market, and involve exclusivity conditions or the unavailability of the specialization products on a standalone basis[121]. Outside the safe harbor, the Guidelines recognize that an infringement of article 101 by effect is likely to be ruled out in cases where a particular clause is necessary for the parties to engage in joint production in the first place, for instance because it allows for the emergence of a new product or service[122]. More generally, however, it will be necessary to take into account the specific market conditions: in particular, the market share of the parties, the degree of competition between them, demand and supply substitability, and whether one of the parties is an important competitive force[123]. Particularly illustrative and relevant for data collaboratives is the discussion on collusion achieved through commonality of costs: the Guidelines clarify that this is a concern if production costs constitute a large portion of variable costs and if the market characteristics are conducive to collusion, especially where undertakings already had significant commonality of costs prior to the agreement[124]. In the data pooling scenario mentioned above, this could imply that society stands to lose from this kind of agreements to the extent that the concerned undertakings would have brought to the market, in the absence of the agreement, a higher variety or a more innovative type of insights. Ultimately, the outcome will depend on balancing the effects on different parameters of competition.

Purchasing agreements

A fourth category addressed by the Guidelines, and which is of particular importance for data cooperatives, is that of agreements concerning the joint purchasing of products. Much like production agreements, purchasing agreements can be structured in different forms: they can be done through a jointly controlled company, through a company in which many other companies hold non-controlling stakes, or even by individual contractual arrangement and

[120] *Id*., para. 150
[121] Commission Regulation (EU) No 1218/2010 of 14 December 2010 on the application of Article 101(3) of the Treaty on the Functioning of the European Union to certain categories of specialisation agreements Text with EEA relevance, OJ L 335, 18.12.2010, p. 43–47, Art. 2
[122] Horizontal Cooperation Guidelines, supra n. 70, para. 160 and 163
[123] *Id*., para. 173
[124] *Id*., para. 177-180

looser forms of co-operation[125]. While they are generally viewed as beneficial, since the creation of buyer power can lead to lower prices and better quality to consumers, they may also raise competitive concerns: this is the case if the parties to the agreement have market power either in the purchasing market (such that they may force suppliers to reduce the range or quality of products) or in the selling market (in which case they may soften competition between them and not pass on to consumers any gains from lower purchase prices). If the purchasers are not competing in a relevant market, however (for example because of geographic factors), then the anticompetitive effects of these agreements are unlikely, unless their respective position in the purchasing market is likely to be used to harm the competitive position of other players in that market[126]. It is not entirely clear what the Commission meant by this reference to competitive position of "other players": perhaps it is a generic reference to competitors, in which case, it is suggesting that companies cannot take advantage of the scale achieved through buyer alliances to gain a competitive edge *vis a vis* other resellers; or perhaps it is meant to protect input producers in those markets, preventing buyer arbitrage. In any case, this ambiguity partly frustrates the Commission´s attempt to provide legal certainty.

More generally, the challenge for competition analysis is how to estimate market power in the context of a data collaborative solution: this is relevant both where a collective trades data from its members in exchange for a product or service (let us refer once again to diagnostic data to develop algorithmic insights, for instance in car manufacturing) and where data is the object of the joint purchase (for instance, buying medical prescription data from pharmacies in order to predict trends and optimize production). In both cases, it may be misleading to rely on the traditional approach, especially on the buyer side, as this may lead to underestimate the number of potential buyers.

Standardization agreements

A last category that is relevant for our analysis is that of agreements that have as their primary objective the definition of technical or quality requirements with which current or future products, production processes, services or methods may comply[127]. In addition to what may be called "standardization agreements" *tout court*, which determine technical and quality standards for future products, the Guidelines cover the use of standard terms and conditions for actual products: in particular, the sale or purchase of goods and services between competitors and consumers for substitute products[128]. The Guidelines set out a general framework of assessment, which is based on the general application of a "by object" infringement of article 101 TFEU to agreements that are aimed at excluding actual or potential competitors from the market, or that directly influence prices[129], and an evaluation of the effects of the agreements in other cases[130]. Right away, then, it can be concluded that any agreements involving uniform pricing for participants to a data space will fall foul of article 101 (1). This does not rule out the applicability of an efficiency defense pursuant to article 101 (3), but the conditions for the application of the defense are stringent. The Guidelines identify possible efficiencies, including savings in transaction costs and facilitation of entry, and even establish a positive presumption

[125] *Id*., para. 194
[126] *Id*., para. 212
[127] *Id*., para. 257
[128] *Id*., para. 270-272
[129] in the case of standardization agreements tout court, specifically through disclosure of the most restrictive licensing terms as a cover to jointly fix prices, either for the final product sold on the market or of the substitute IP rights or technology. See *id.,* para. 273
[130] including cases that involve *ex ante* disclosures of most restrictive licensing terms that are not simply a cover to fix prices. See *id*., para. 299

for standards that facilitate technical interoperability and compatibility[131]. This can be useful for initiatives that set common standards for the transfer and re-use of datasets, which is a very important component of data collaboratives. However, it is plausible that this presumption would be self-defeating if the enhanced interoperability or compatibility is exclusive, in particular where it favors an entrenched group of competitors who collectively have market power. That presumption is likely to bear more wight where the effect is not merely vertical, but also horizontal[132], meaning that it stimulates competition not only within an ecosystem but across different ecosystems.

To guide the "by effects" analysis, the Guidelines provide a broad safe harbor- the only one that is not linked to a market share threshold- for standardization agreements *tout court* that respect four cumulative conditions: (a) an unrestricted industry participation in a transparent standard-setting procedure; (b) the inexistence of any obligation to comply with the adopted standard; (c) good faith disclosure of standard-essential intellectual property rights; and (d) a clear and balanced IP rights policy, including accessibility to the standard on fair, reasonable and non-discriminatory (so called "FRAND") terms. A similar, but adjusted version of the safe harbor is also provided for the use of standard terms, by replicating conditions (a) and (b) above and adding the requirement of effective availability for everyone (without specifically linking this to the use of FRAND terms) [133]. Thus, in comparison to the safe harbor for standardization agreements *tout court*, there is no requirement involving intellectual property (on the assumption that its use as an input for future products is not relevant to these agreements) and no explicit transparency requirement the procedures leading to the adoption of the standard terms[134]. Furthermore, two important exceptions are made to this safe harbor, identifying situations in which it will be necessary to examine the effects of these agreements: first, when their terms define the scope of the product sold to consumers[135]; and second, when they become *de facto* standards in the industry[136]. The first exception can be explained noting that standard terms imposing a limit on future production have effects comparable to those of standard-setting, in keeping out of the market some products for which there would have been potential consumer demand. Similarly, the second exception can be understood considering that *de facto* standardization can have effects that are equivalent to the imposition of binding standards, as it results in wide market adoption.

In the context of the data economy, both technical standardization and the use of standard terms are likely to be crucial factors in stimulating data portability and interoperability, delivering substantial benefits in terms of data use and analysis[137]. This begs the question of whether the safe harbors we have just seen would be applicable, *mutatis mutandis*, to this different environment: in particular, it is plausible that standardization agreements generate anti-competitive effects regardless of their respect of the four aforementioned conditions, to the extent that the setting of interoperability and portability standards is done by undertakings with market power (and without the need to establish that they amount to *de facto* standards) and

[131] Para. 321
[132] Heike Schweitzer and Wolfgang Kerber, "Interoperability in the digital economy" (Macie Paper Series, Nr. 2017/02, 2017) <https://econpapers.repec.org/paper/marmagkse/201712.htm> accessed 03 March 2021

[133] Horizontal Cooperation Guidelines, supra n. 70, para. 272 and 301
[134] *Id.,* para. 280
[135] *Id*. Para. 303
[136] *Id.*, para. 305
[137] Gal, Michal and Rubinfeld, Daniel L., Data Standardization (June 2019). 94 NYU Law Review (2019) Forthcoming, NYU Law and Economics Research Paper No. 19-17, Available at SSRN: https://ssrn.com/abstract=3326377 or http://dx.doi.org/10.2139/ssrn.3326377

that these undertakings manage to control or substantially influence the standard-setting process[138]. This doubt is reinforced by the soft wording of the two safe harbors, which refer to agreements that "normally" will fall outside the scope of article 101[139] and that are not likely to give rise to restrictive effects on competition[140], thus not leaving prospective co-operators with much certainty.

3.2. The competition implications of public-private partnerships

In addition to the interpretative challenges raised by data collaboratives from the perspective of private agreements, a further layer of complexity concerns their interaction with public power. In the European Union, a Member State can be held responsible for a violation of the competition rules if it requires or favours the adoption of agreements, decisions or concerted practices contrary to Article 101, reinforces their effects, or deprives its own legislation of its official character by delegating to private traders responsibility for taking decisions affecting the economic sphere[141]. This responsibility, which is jointly shared by the undertakings unless they have been strictly required to take such measures (without residual discretion) arises from the duty of sincere cooperation in article 4 (3) of the Treaty of the European Union, which requires Member States to facilitate the achievement of the Union's tasks and refrain from any measure which could jeopardise the attainment of the Union's objective[142]. It therefore should be considered for national implementations of data collaboratives, and potentially also for EU data spaces- to the extent that there is any need to involve national legislators for the establishment or continued operation of the data spaces.

Article 4 (3) could potentially also be a ground for responsibilization of the EU legislature or the EU Commission in the implementation of legislation, as long as their action has a bearing on national measures to be subsequently adopted: the article creates a duty of sincere cooperation for EU institutions any interactions with national legislators, courts or other institutions necessary to give full effect to EU law[143]. Furthermore, even fully autonomous EU legislation could be held invalid in case of violation of rules relating to the application of the EU treaties, including articles 101 and 102 TFEU, if brought within the strict time limits[144] or if the European Court of Justice ascertained the incompatibility with the EU´s primary law within the context of a preliminary reference proceeding (as occurred for instance in *Digital Rights Ireland*[145]). However, the important difference between EU and State responsibility is that a State may be found jointly liable with any concerned undertakings for a violation of the competition rules that has been facilitated by a State measure, while the responsibility may

[138] Nicolo Zingales and Olia Kanevskaia, The IEEE-SA patent policy update under the lens of EU competition law. European Competition Journal 12 (2-3), 195-235, 2016
[139] Horizontal Cooperation Guidelines, supra n. 70, para. 278
[140] *Id.,* para. 301
[141] *Pascal Van Eycke v Société anonyme ASPA* (1988) Case 267/86, ECLI:EU:C:1988:427, at 17
[142] Consolidated version of the Treaty on European Union, OJ C 326 (2012), Art. 4(3)
[143] E.g. Case 120/73, *Lorenz v. Germany*, 1973 ECR 1471; Case C-354/90, *FJdJration Nationale*, 1991 ECR I 5505 at paras. 9-11. See in this sense also the speech by the Director General of Competition, John Temple Lang, "The core of the constitutional law of the Community - Article 5, EC Treaty" (1995), available at https://ec.europa.eu/competition/speeches/text/sp1995_024_en.html#foot68
[144] Consolidated version of the Treaty on the Functioning of the European Union, OJ C 326 (2012), Art. 263
[145] Joined Cases C-293/12 and 594/12, *Digital Rights Ireland and Seitlinger and others*, ECLI:EU:C:2014:238

never be shared with EU institutions, regardless of a presumed incompatibility of their measures with the competition rules[146].

The above suggests that there is more leeway in pursuing legislative and policy options for data spaces at the EU level, as this would create a favourable environment for collaborations without the risk of triggering liability for the undertakings concerned. However, this strategic choice may not bode well with the division of competences envisaged by the EU Treaty, and in particular the principles of conferral, subsidiarity and proportionality. The limit set by the principle of conferral is that the EU only has the competences that are attributed to it by Member States in its Treaties, while the principle of subsidiarity allows EU action in areas in which the Union does not have exclusive competence "only if and in so far as the objectives of the proposed action cannot be sufficiently achieved by the Member States, either at central level or at a regional and local level, but can rather, by reason of the scale or effects of the proposed action, be better achieved at Union level"[147]. Furthermore, even where there is scope for EU action in this sense, it must be confined within the limits set by the proportionality principle, according to which "the content and form of Union action shall not exceed what is necessary to achieve the objectives of the Treaties"[148]. On the basis of these principles, one may question the extent to which the EU would be acting fully within its powers when establishing the conditions under which members of data collaboratives ought to organize themselves, including their establishment and their ordinary administration. The necessity of EU action can be doubted, in particular, where the EU is acting in an area of shared competence with EU Member States, where the latter may have developed a policy or legislation that may be in tension with the one pursued by the Union[149].

Let us make the controversy a little more concrete by hypothesizing a tension between the national rules for data protection law, and an action taken by the EU to create an EU data space. Despite the harmonizing role of the GDPR, significant margin is left for EU Member States in the implementation of a number of its provisions: indeed, a recent report for the European Commission documented wide divergence between States on the implementation of some key provisions in the Regulation, notably articles Article 8(1), Article 9(4), Article 23(1), points (c) and (e), Article 23(2), Article 85(1) and (2) and Article 89(2), (3) and (4) GDPR.[150] .

[146] Where such incompatibility has been established by the European judicature, the concurrent responsibility with the EU institution cannot be invoked, but anyone who has suffered consequences could in principle bring an action for non-contractual liability against the EU institution concerned. See Art 340 TFEU.

[147] Consolidated version of the Treaty on European Union, OJ C 326 (2012), Art. 5 (3)

[148] *Id.,* 5 (4)

[149] According to article 2 (2) TFEU: "When the Treaties confer on the Union a competence shared with the Member States in a specific area, the Union and the Member States may legislate and adopt legally binding acts in that area. The Member States shall exercise their competence to the extent that the Union has not exercised its competence. The Member States shall again exercise their competence to the extent that the Union has decided to cease exercising its competence."

[150] European Commission, Report on the implementation of specific provisions of Regulation (EU) 2016/679. To illustrate, Article 8(1) GDPR concerns the age that a child must be in order for his consent to be valid for the processing of his personal data in relation to information society services. Article 9(4) GDPR allows Member States to maintain or introduce further conditions, including limitations, with regard to the processing of genetic data, biometric data or data concerning health. Article 23(1)(c) and (e) GDPR permits Union and national legislation to restrict the scope of the obligations and rights provided for in Articles 5, 12 to 22 and Article 34 GDPR, where such restrictions respect the fundamental rights and are necessary and proportionate to safeguard public security and other important objectives of general public interest. Article 23(2) GDPR stipulates a number of accompanying conditions and safeguards, when restrictions under Article 23(1)(c) and (e) GDPR apply. Article 85(1) GDPR requires Member States to reconcile the right to the protection of personal data with the right to freedom of expression and information, while Article 85(2) GDPR requires Member States to provide, where appropriate, for exemptions or derogations from Chapters II to VII and IX GDPR for processing carried out for

Diverging rules could relate, for instance, to the conditions for processing biometric data, and the proper balancing between the protection of personal data and freedom of expression, or between processing of personal data and scientific or historical research purposes, statistical purposes or achieving purposes in the public interests. *Quid iuris*, then, for an official EU act which potentially overrides the additional protections afforded by national legislation of this type?

The Union´s measure will be scrutinized from a subsidiarity and proportionality standpoint, in the context of which particular regard will be given to the legal basis upon which the Union justifies its action, and the relevant interest that is declared to be pursued. For instance, in the case of the proposal for a Data Governance Act that was introduced in December 2020[151] as a first package of the measures announced in the European Data Strategy, the general overarching objective is to leverage the potential of data for the economy and society; and the more specific objectives to reinforce trust in data sharing, to make more public sector data available for reuse and facilitating the collection of data to be used for the common good, and to improve data findability, data quality and data interoperability across sectors and countries[152]. The need for EU action in this case is premised upon the recognition that the key sectors of the economy span across borders (with suppliers, producers and clients established in different Member States), that companies active in the data economy should be able to benefit from the size of the internal market by rolling out EU-wide products and services, and that there is a risk of legal and administrative fragmentation without an EU act that lays down the elements that ensure comparable access and use conditions in all data spaces[153]. Compared to national legislation, EU action would give the added benefit of promoting mutual recognition of certification/labelling mechanisms and of a trust scheme for data altruism, and of an ability of companies specialized in data sharing services to offer these in all Member States[154].

Having listed these goals and challenges, the Commission identified as the legal basis for its action in the Data Governance Act article 114 TFEU – which allows the Union to take measures having as their object the establishment and functioning of the internal market. The same legal basis (and essentially the same factual justifications) is declared also in the impact assessment for the Data Act, the second piece of legislation of the European Strategy for Data, which has as its primary objective facilitating business to business data sharing[155]. However, it should be noted that there are limits to the use of this legal basis, which does not enjoy the same kind of legitimacy of the more permissive legal basis of article 352 TFEU (for EU action that proves necessary to attain one of the objectives set out in the Treaties): it does not require unanimity voting at the Council, and therefore can be taken by the EU without the support of all Member States. For this reason, article 114 TFEU does not grant a general right to regulate: it needs to be clearly shown that differences among Member States are such as to obstruct the fundamental freedoms, and thus has a direct effect on the establishment and functioning of the internal

journalistic purposes or the purpose of academic, artistic or literary expression. Finally, Article 89(2), (3) and (4) GDPR provides Member States the option to derogate from certain data subject rights where personal data are processed for scientific or historical research purposes, or statistical purposes, or archiving purposes in the public interest, subject to certain conditions and safeguard

[151] Proposal for a Regulation of the European Parliament and of the Council on European data governance (Data Governance Act) (2020), COM/2020/767

[152] Impact Assessment for the Data Governance Act, pp. 18-20

[153] Data Governance Act (DGA), supra n. 151, p. 2

[154] *Id*., p. 6

[155] Inception Impact Assessment of EU Data Act, Ref. Ares (2021)3527151 - 28/05/2021

market[156]. In this regard, the CJEU has clarified that the choice of the legal basis, and thus, the existence any future likely fragmentation in the absence of the measure, may not depend simply on an institution's conviction as to the objective pursued: it must be based on objective factors amenable to judicial review[157]. It will thus be necessary to establish objective elements illustrating why such fragmentation is likely to arise, in addition to showing why the measures at stake are suitable and proportionate to address the risks arising from fragmentation.

When we examine the provisions introduced by the proposed Data Governance Act and the types of interventions envisaged in the forthcoming Data Act, however, it is not clear that these conditions are consistently met. For instance, it is not clear that the evidence of likely future fragmentation is sufficient to justify all three building blocks of the Data Governance Act, namely (i) the conditions for the re-use, within the Union, of certain categories of data held by public sector bodies; (ii) a notification and supervisory framework for the provision of data sharing services; and (iii) a framework for voluntary registration of entities which collect and process data made available for altruistic purposes. In particular, doubts can be casted on the necessity of each of the detailed list of conditions imposed to data sharing service providers (including neutrality and fiduciary duties)[158], as well as the need to establish a new regulatory regime for data sharing services and data altruism, and to introduce competent authorities to administer these regimes[159]. Similarly, within the context of the forthcoming proposal for a Data Act, one could question the extent to which national divergences in the law of unfair commercial practices or trade secrets law ought to justify a comprehensive set of measures and regulatory arrangements designed to facilitate data access and use, both within the public and the private sector. At the same time, even if the use of article 114 as legal basis were to be accepted, we should not forget that a mechanism exists that enables the Commission to take into account conflicting interest of Member States on grounds of major needs referred to in Article 36 (non-economic considerations such as public morality, public policy or public security) or relating to the protection of the environment or the working environment, and thereby authorize that State to introduce a national provision which derogate from a harmonization measure[160]. Notably, this may be the case because the data space is physically stored in one Member State´s jurisdiction, or because it involves data about its citizens. As a

[156] Case C-376/98, *Federal Republic of Germany v European Parliament and Council of the European Union* (*Tobacco Advertising I*), ECR 2000 I-08419 paras 84 and 95, *Arnold André* (n 2), para 30; Case C-380/03 *Federal Republic of Germany v European Parliament and Council of the European Union* (*Tobacco Advertising II*), ECR 2006 I-11573, para 37; C-151/17, *Swedish Match AB contra Secretary of State for Health*, ECLI:EU:C:2018:938, para 29; Case C-491/01 *British American Tobacco (Investments) and Imperial Tobacco "BAT"* EU:C:2002:741, paras 59 and 60 and Case C-58/08 *Vodafone and Others* EU:C:2010:321, para 32.

[157] Case 45/86, *Commission v Council*, ECLI:EU:C:1987:163, para. 11.

[158] DGA, supra n. 151, Chapter III

[159] *Id.,* Art. 7, 12, 23-29

[160] In particular, art 114 (4) and (5) provide that: If, after the adoption of a harmonisation measure by the European Parliament and the Council, by the Council or by the Commission, a Member State deems it necessary to maintain national provisions on grounds of major needs referred to in Article 36, or relating to the protection of the environment or the working environment, it shall notify the Commission of these provisions as well as the grounds for maintaining them" and that "[…] without prejudice to paragraph 4, if, after the adoption of a harmonisation measure by the European Parliament and the Council, by the Council or by the Commission, a Member State deems it necessary to introduce national provisions based on new scientific evidence relating to the protection of the environment or the working environment on grounds of a problem specific to that Member State arising after the adoption of the harmonisation measure, it shall notify the Commission of the envisaged provisions as well as the grounds for introducing them." As per art 114 (6) and (7), these notifications trigger the duty for the Commission to accept or reject the provision(s) involved, after having verified whether or not they are a means of arbitrary discrimination or a disguised restriction on trade between Member States and whether or not they shall constitute an obstacle to the functioning of the internal market, and the possibility for the Commission to adapt the harmonizing measure accordingly.

result, the possibility that Member States play a significant role in the design and implementation of the rules applicable to data spaces should not be discarded, at least until the adoption of the aforementioned legislation.

This is where the potential competition problems arising from the uncertainty in the assessment of data collaboratives come back to bite: as discussed above, the involvement of a Member State makes a crucial difference if the measure it takes favors the adoption of agreements, decisions or concerted practices contrary to Article 101, reinforces their effects, or deprives its own legislation of its official character by delegating to private traders responsibility for taking decisions affecting the economic sphere. In these circumstances, not only the State, but also the undertakings concerned will be found in violation of said article. As a matter of policy, it would have been strategic for the European Union to limit the scope for Member States intervention, so as to prevent this from acting as a disincentive against joining an EU data space. However, as discussed below, there are a couple of provisions in existing and proposed EU legislation that could potentially lead to this outcome at the public-private interface: these are specific provisions in the Data Governance Act (DGA) and the Directive on open data and the re-use of public sector information (PSI Directive)[161].

The PSI Directive is an important piece of legislation determining legal obligations relating to the re-use of documents produced by State, regional or local authorities and so-called "public sector bodies": bodies that are established for the specific purpose of meeting needs in the general interest (thus not having industrial or commercial character) and that are financed, for the most part by the State, regional or local authorities, or by other bodies governed by public law; or are subject to management supervision by those authorities or bodies; or have an administrative, managerial or supervisory board, more than half of whose members are appointed by the State, regional or local authorities, or by other bodies governed by public law[162]. With regard to its scope, this Directive marks a significant departure from its predecessor[163], by comprising within its regime so-called "public undertakings", i.e. undertakings active in the areas of public functions (e.g. utlilities, railway, airlines, shipping) over which the public sector bodies may exercise directly or indirectly a dominant influence by virtue of ownership, financial participation or the rules which govern them[164]. However, there is no general obligation for such undertakings to permit re-use of their documents, and it will be only to the extent that such re-use is permitted that they will need to follow the relevant provisions of the Directive, in particular as regards format, charging, transparency, licences, non-discrimination and prohibition of exclusive arrangements[165]. It should also be clarified that a number of documents are excluded from the regime, including those held by public undertakings that have been produced outside the scope of the provision of services in the general interest as defined by law or other binding rules in the Member State; those of public undertakings that are related to activities directly exposed to competition, and therefore not subject to procurement rules; those the supply of which is falling outside the scope of the public task of the public sector bodies concerned; those for which third parties hold intellectual property rights; and those which are excluded from access by virtue of the access regimes in the Member State concerned, including on grounds of the protection of national security

[161] Directive (EU) 2019/1024 of the European Parliament and of the Council on open data and the re-use of public sector information (2019), OJ L 172
[162] Id., art. 2 (2)
[163] Directive 2003/98/EC of the European Parliament and of the Council of 17 November 2003 on the re-use of public sector information. OJ L 345, 31.12.2003, p. 90–96
[164] PSI Directive, supra n. 161, art. 2 (3)
[165] *Id.,* Recital 26

(namely, State security), defence, or public security, statistical confidentiality, commercial confidentiality (including business, professional or company secrets), the protection of personal data and the protection of sensitive critical infrastructure as defined in point (d) of Article 2 of Directive 2008/114/EC[166].

These carve-outs are particularly significant in the context of data spaces, as much of the data that is intended to be shared may have been industrially produced (and therefore involving intellectual property rights) or of sensitive nature (for instance because the data is of personal or confidential nature). The DGA fills a gap in this respect, as it specifically applies to data held by public sector bodies which are protected on grounds of commercial confidentiality, statistical confidentiality, the protection of intellectual property rights of third parties and the protection of personal data[167], although on the other hand it excludes from its application data held by public undertakings and some other specific public sector bodies (such as public service broadcasters and cultural establishments)[168]. Another crucial difference between the two legislations is that the latter does not impose an obligation to allow re-use, but simply leaves that determination to the public sector body concerned (absent specific obligations arising from EU or national law)[169], similar to the regime imposed by the Directive on public undertakings. However, to the extent that re-use is allowed, the public sector body shall make publicly available the conditions regarding re-use, which must be non-discriminatory, objective and justified, and not be used to restrict competition.[170] Possible conditions include the anonymization or pseudnonimization of personal data or the deletion of commercially confidential information, restrictions necessary to ensure secure processing, obtaining consent from data subjects or permission from the legal entities that may be adversely affected, accessing and re-suing the data within a secure processing environment provided and controlled by the public sector, and accessing and re-using the data within the physical premises in which the secure processing environment is located (in particular if remote access cannot be allowed without jeopardising the rights and interests of third parties)[171].

Where the two legislations reveal complementarity is with regard to two key aspects in the re-use of documents held by public sector bodies: charging and exclusivity. With regard to charges, the PSI Directive establishes that there must be none for the re-use of documents, but that this is without prejudice to the possible compensation of the marginal costs incurred for the reproduction, provision and dissemination of documents as well as for anonymisation of personal data and measures taken to protect commercially confidential information. The same article carves out three exceptions to the gratuity of re-use, namely for (a) public sector bodies that are required to generate revenue to cover a substantial part of their costs relating to the performance of their public tasks; (b) libraries, including university libraries, museums and archives; and (c) public undertakings. Even in such cases, however, the total income derived from supplying and allowing the re-use of documents shall not exceed a reasonable return on investment combined with the cost of collection, production, reproduction, dissemination, data storage (and in the case of libraries, museums and archives, also the preservation and rights

[166] *Id.,* Article 1(2)
[167] *Id.,* Art 3 (1)
[168] Specifically, data held by public service broadcasters and their subsidiaries, and by other bodies or their subsidiaries for the fulfilment of a public service broadcasting remit; data held by cultural establishments and educational establishments; data protected for reasons of national security, defence or public security; data the supply of which is an activity falling outside the scope of the public task of the public sector bodies concerned. *Id.,* art. 3 (2)
[169] *Id.,* art. 3 (3)
[170] *Id.,* art. 5 (1) and (2)
[171] *Id.,* art. 5 (2)- (6)

clearance), anonymisation of personal data and any measures taken to protect commercially confidential information; and in the case of (a) and (c), the total charges shall be calculated in accordance with objective, transparent and verifiable criteria laid down by Member States. This is quite important for our purposes, as allowing Member States to dictate conditions for the calculation of these costs may directly or indirectly lead to uniformity in the pricing schemes of data spaces within a specific jurisdiction, and to different equilibria across jurisdictions. While the fact that the fees would be linked to some specific measures of cost limits the possible ways in which prices could be inflated, this does not rule out the possibility that those factors generate anticompetitive effects[172]: for instance, because they restrict some specific parameter of competition by imposing a cap on the amount of expenses related to anonymization that can be subject to compensation. This effect may be reinforced by the transparency requirements imposed on public sector bodies, including the obligation to publish the amount and calculation basis for any standard charges, to indicate at the outset any factors taken into account in the calculation of non-standard charges[173].
The DGA contains a similar provision is introduced that, while allowing the charging of fees in exchange for the re-use of data, prescribes that such fees shall be derived from the costs related to the processing of requests for re-use and based on a methodology to be published in advance, along with a description of the main categories of costs and the rules used for the allocation of costs[174]. However, unlike the obligation imposed on public sector bodies to impose[175] and make publicly available[176] the conditions regarding re-use, this provision does not include a *chapeau* (similar to the one applicable to the use of public policy justifications under article 36 TFEU) to ensure that these conditions are not used as a way to limit competition. Thus, the proposed DGA does not prevent Member States from what effectively amounts to regulating (part of) the pricing of data spaces, even where this may lead to collusion.

As to exclusivity, both legislations aim to tackle an important barrier to the re-use of public sector information: agreements or other arrangements between public bodies and third parties which grant to the latter exclusive rights in relation to the documents held by the former. While each contains general prohibitions of exclusivity, they preserve the possibility to do so when that is necessary for the provision of a service or a product that is of public general interest, within appropriate limits. In particular, the PSI Directive requires that the reason for granting such an exclusive right shall be transparent and made public, and subject to regular review (at least every three years). Similarly, the DGA requires that any period of exclusivity does not exceed three years, and imposes compliance with the principles of transparency, equal treatment and non-discrimination on grounds of nationality[177] However, only the PSI Directive includes provisions targeting legal and practical arrangements that, short of granting exclusivity, lead to a restricted availability for the re-use of documents other than the third party participating in the agreement: these *de facto* exclusivity arrangements shall be made publicly available online at least two months before their coming into effect, be subject to regular reviews (at least every three years) as to their effect on the availability of data re-use, and be transparent and publicly available online[178]. The fact that the DGA does not have equivalent provisions means that, when it comes to documents burdened by intellectual

[172] See, by analogy, the arguments advanced to challenge the legality of the pricing methodology introduced by the IEEE Standards Association with its patent policy update in 2015: Nicolo Zingales and Olia Kanevskaia, supra n. 138.
[173] PSI Directive, supra n. 161, Art 8
[174] *Id.* Art. 6
[175] *Id.* Art. 8 (1) PSI
[176] Proposed Data Governance Act, supra n.151, Art 5 (2)
[177] *Id.,* Art 4 (4)- (6)
[178] PSI Directive, supra n. 161, art. 12

property rights or other exceptions to the general access regime, there is no mechanism to prevent arrangements made between public and private entities establishing conditions for access that restrict availability for market participants other than the beneficiary in ways that are different from *de iure* exclusivity. This loophole appears to allow a Member State not only to reinforce the effects of anticompetitive agreements made between the beneficiary and other undertakings, but also to favor the adoption of such agreements, or else to delegate important decisions in the hand of the private sector, all of which would give rise to a joint responsibility of that Member State and any concerned undertaking for a violation of article 101 TFEU. For instance, a health authority could design an access and re-use regime that heavily depends on a complex architecture for anonymization and consent management which is only available for some more established industry players, thereby potentially undermining the goal of unlocking the potential of data re-use within the internal market.

4. A problem with a viable solution: innovation hubs and regulatory sandboxes

The above analysis has demonstrated the complexity of competition law analysis in the context of data collaboratives, and EU data spaces in particular. This was done in three steps.

First, in section 1, it was shown that data collaboratives can take different forms, depending on the aspirations and constraints of their members. In the most extreme case, known as "data union", individuals get together to represent the interests of a community, and coordinate on the use of their data in order to improve their socio-economic conditions. A similar, but more malleable type of organization is that of data cooperatives, which represent an autonomous association of persons united voluntarily to meet their common economic, social, and cultural needs and aspirations through a jointly-owned and democratically-owned enterprise. The concept of data trust can also be aligned with the common economic, social and cultural needs of their members, but presupposes the delegation of key decisions to an independent third party. Finally, the concepts of data exchanges and data pools place more emphasis on the data that is made available, rather than the internal arrangements that underpin relationships between their members: they are market-driven initiatives, which may include data as well as technology or infrastructure.

Second, Section 2 discussed the concept of "EU data space" promoted by the European Commission in a Communication from 2020, where it lays out its vision to "become a leading role model for a society empowered by data to make better decisions". To reach that goal, the Communication outlines a strategy for policy measures that include the establishment of effective rules to ensure trustworthy technologies and the creation of various thematic "data spaces" -a term not specifically defined, but which evokes the idea of repositories for the secure and interoperable sharing of data. It was noted that the Commission places great emphasis on the legally compliant nature of collaborations occurring within these data spaces, without however detailing how such compliance would be ensured. Relatedly, a doubt was casted on the ability of these data spaces to be sufficiently attractive for smaller industry players, particularly in the absence of contribution by their more data-reach competitors. It was concluded that the devil is in the detail, and thus much will depend on the ways in which data spaces will be structured.

Third, Section 3 approached the question of compliance from a competition law perspective. Competition law analysis has a number of guidelines that can inform its application in the context of so-called "horizontal cooperation" agreements. It was noted that the guidance and

case-law do not really enable undertakings to predict the way in which information exchanges will be evaluated, especially where they could bring benefits for only a subset of consumers. Similarly, analyzing data collaboratives as R&D agreement raises more questions than answers, due to the instrumental role of data as input for products in new markets, and the difficulty of defining whether datasets are complementary or substitutes. Production agreements may be relevant to the extent that data holders collaborate for the logistical production of a particular product, and to that end, the assessment of legality will depend on the specific market conditions: in particular, the market share of the parties, the degree of competition between them, demand and supply substitability, and whether one of the parties is an important competitive force. The Guidance for purchasing agreements goes one step further, by explaining that a purchasing cooperative may be justified if that coordination was necessary to obtain a particular product or service, to the extent that the parties do not have market power. However, when it comes to data collaboratives it is more challenging to verify the applicability of this last condition: case-law and commentators are far from univocal in identifying the criteria that are indicative of market power in relation to data. Finally, the Guidelines sketch a safe harbor framework for standardization agreements, but it is unclear to which extent this is applicable to conduct like the standard license of one or more data collaboratives, calling for a casuistic evaluation of the circumstances.
In its document presenting the EU Data Strategy, the European Commission remained silent on these interpretative challenges, giving the impression that somehow the interpretative challenges would disappear after formalizing the appropriate structure for each data space. However, it was pointed out that the formalization presents its challenges too, as the prospect of Member State´s responsibility for the drafting of the rules may indicate that there is a violation of competition rules both by that State and the undertaking concerned. This is of particular concern when it comes to the rules on pricing and exclusivity for the re-use of documents held by public sector bodies and public undertakings, since they both preserve the space for Member States to establish terms of general application, and these can be found to have anti-competitive effects. Furthermore, given the pivotal role played by data to enter new markets and improve existing products and services, the concern of compliance with data protection rules becomes paramount in this context.

Given the challenges outlined above, we propose a solution borrowed from financial regulation that would ameliorate many of the concerns of legal certainty and regulatory arbitrage that have been discussed: the use of innovation hubs and regulatory sandboxes to steer data collaboratives towards the best form of cooperation agreements, and to facilitate compliance with the applicable regulation. According to a comprehensive report recently published by the European Banking Authority, the European Securities and Market Authority and the European Insurance and Occupational Safety Authority[179], an innovation hub provides a dedicated point of contact for firms to raise enquiries with competent authorities, and to seek non-binding guidance on the conformity of their products and services with the applicable regulatory framework. By contrast, a regulatory sandbox provides a scheme to enable firms to test, pursuant to a specific testing plan agreed and monitored by a dedicated function of the competent authority, innovative products, services or business models, typically. This may also imply the use of legally provided discretions by the relevant supervisory authority, which

[179] ESMA, EBA, EIOPA, Report: FinTech: Regulatory sandboxes and innovation hubs. JC 2018 74. Available at <https://www.eba.europa.eu/sites/default/documents/files/documents/10180/2545547/154a7ccb-06de-4514-a1e3-0d063b5edb46/JC%202018%2074%20Joint%20Report%20on%20Regulatory%20Sandboxes%20and%20Innovation%20Hubs.pdf > (accessed 1 June 2021)

typically is one of national level, considering that these experiments have been conducted in the financial sector.

The concept of innovation hub is relatively straightforward and easy to implement: all it requires is the establishment of a contact point with regulators where undertakings can raise questions, and possibly initiating a dialogue, about the application of regulatory and supervisory requirements to innovative business models. This is equivalent to the so-called "comfort letters" which have been specifically recognized as an administrative action available for the European Commission[180], but have never been used so far[181]. The use of this mechanism for guidance on antitrust matters has also been suggested by the Commission within its Communication on the temporary framework to assess business cooperation in response to COVID-19[182]. However, the challenge involved in providing an adequate innovation hub for data spaces would require the involvement of other regulators, too (for instance, data protection authorities and any relevant thematic regulator for the data space in question), thus requiring the signature of inter-institutional cooperation agreements to enable joint regulatory responses. This is not entirely novel to innovation hubs, as in some countries MoUs have been signed even within the financial sector in order to allow the coordination between central banks and financial market authorities[183].

The more provocative proposition made in this chapter, however, is the introduction of a "regulatory sandbox" for firms participating in data spaces, which would take the of the innovation hub to the next level. The term sandbox comes from computer science, where it refers to an isolated testing environment that enables users to run programs or execute files without affecting the application, system or platform on which they run.[184] The importation of this concept into the regulatory environment began in 2015, when the financial markets regulator of the United Kingdom (Financial Conduct Authority) launched this experiment to promote competition and innovation in fintechs. Since then, sandboxes have mushroomed around Europe, and not only in the financial sector: for instance, AI sandboxes have been recently launched by the data protection authorities in the United Kingdom, Norway and France[185].

The typical process for a sandbox is the following: the regulator issues a call for applications by undertakings that have important innovations to bring to the market in a specific domain, and candidates will have to make a submission proving that they fulfill the requirements for participation. For instance, the FCA´s sandbox of 2015 required an explanation of the innovation in question, its benefits to consumers and the related business model, the potential risks and how those can be mitigated in the sandbox, why the sandbox would be necessary, and

[180] Commission Notice on informal guidance relating to novel questions concerning Articles 81 and 82 of the EC Treaty that arise in individual cases (guidance letters). Official Journal C 101, 27.04.2004, p. 78-80. Official Journal C 101, 27.04.2004, p. 78-80

[181] European Commission, "Informal guidance" (2020) <https://ec.europa.eu/competition/antitrust/legislation/guidance.html> accessed 01 June 2021

[182] Communication from the Commission Temporary Framework for assessing antitrust issues related to business cooperation in response to situations of urgency stemming from the current COVID-19 outbreak 2020/C 116 I/02 C/2020/3200. OJ C 116I , 8.4.2020, p. 7–10

[183] Specifically, Belgium and Netherlands. See EBA Report, supra n. 179, p. 12

[184] Linda Rosencrance, "sandbox (software testing and security)" (Tech Target, December 2018) <https://searchsecurity.techtarget.com/definition/sandbox> accessed 01 June 2021

[185] Sofia Ranchordas, "Experimental Regulations for AI: Sandboxes for Morals and Mores" (Morals and Machines, vol. 1 (1/2), 2021), <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3839744> accessed 01 June 2021

what the testing plan is[186]. Subsequently, upon approval, selected firms[187] will receive a no-action letter by the regulator with regard to specific obligations that are within the scope of the regulator´s powers. The testing period may vary, depending on the jurisdiction and on the products at stake, but can be interrupted in some circumstances: if the firm fails to comply with the designed testing parameters, if this reveals to be necessary to mitigate consumer detriment and if there is no demand for the proposition, or the proposition fails to work as expected[188].

It is submitted that there are compelling reasons to extend sandboxes to the context of data spaces: first, they would accelerate the creation of this important policy innovation, and thus increase awareness and adoption by market participants in the delicate initial phase of emergence. This is because they would allow providers of the relevant data and infrastructure to put their products under scrutiny of the regulators with the aim to get a better understanding of regulatory requirements, so as to possibly adjust and consolidate their business models. Second, they would offer an elegant way of resolving the interpretative challenges with regard to the enforcement of competition law, by making use of the Commission´s enforcement discretion. This would not technically shield participating undertakings from the risk of private lawsuits, but it would have the effect of allowing experimentation for an initial interim period under close guidance by the European Commission, potentially providing them with a so-called "regulated conduct defence": firms who have merely followed instructions by the Commission could not be held liable for the resulting conduct, as long as they had no discretion in taking a different course of action[189]. Third, the testing would enable the European Commission to collect direct evidence on the working of data space, so as to develop a better understanding of their effects. This is particularly crucial as the properties of certain arrangements may not be intuitive from a first examination: for instance, the possibility that sensitive and personally identifiable information is inferred from the use of common benchmarks[190] or from the sharing of models used for machine learning applications[191] should not been discounted. The ability to examine the effects "on the ground" would be complemented by the filing of reports by participants at the end of the experimentation, linked to the achievement of Key Performance Indicators (KPIs). Fourth, and finally, the European Commission could make use of this mechanism to require the adoption of safeguards not only against the violation of competition law, but also to ensure fair and transparent processing of personal data.

How would it work, in practice, for contributors and/or beneficiaries of data spaces? The sandbox would be selective both with regard to the providers of data or services in data spaces, and with regard to the consumers who can benefit from such services. First, there would be a

---

[186] Financial Conduct Authority, "Regulatory Sandbox" (November 2015), <https://www.fca.org.uk/publication/research/regulatory-sandbox.pdf> accessed 01 June 2021

[187] Though selection may be automatic upon the fulfillment of the criteria: see for instance the case of the Austrian Security and Investment Commission at https://asic.gov.au/for-business/innovation-hub/enhanced-regulatory-sandbox/ accesses 01 June 2021

[188] ESMA, EOIPA & EBA Report, supra n.179, para. 71

[189] See OECD, Policy Roundtable: Regulated Conduct Defence (2011), available at <https://www.oecd.org/regreform/sectors/48606639.pdf> accesses 01 June 2021

[190] A recently opened antitrust investigation on this issue involves the US poultry industry´s use of a common platform (Agristat) to share statistics about farmer pay, the amount of time elapsed between a farmers' flocks and how farmer compensation in one region compares to the average for all farmers across the country. According to reports, poultry companies could easily reverse engineer the identity of profiled farmers, which in turn led to collusion. See <https://www.theguardian.com/environment/2019/aug/03/is-the-us-chicken-industry-cheating-its-farmers> accessed 1 June 2021

[191] Veale Michael, Binns Reuben and Edwards Lilian 2018Algorithms that remember: model inversion attacks and data protection lawPhil. Trans. R. Soc. A.3762018008320180083 http://doi.org/10.1098/rsta.2018.0083

screening as to the data that can be inputted in a data space and the undertakings that can participate to it. The European Commission could issue a call specifying minimum standards for admission to the data space, including guarantees of financial stability, antitrust compliance measures and a number of requirements relating to data processing, such as the identification of data provenance, the use of state-of-the-art security measures and a clear allocation of responsibilities for the data processing within the data space. At this stage, the Commission could filter out from the testing datasets of dubious quality or which may give rise to significant liability risks, from a data protection, cybersecurity and competition standpoint: indeed, the compliance with these requirements could be facilitated by cooperation with the national competent authorities envisaged by the DGA, which have the power to exchange information with the data protection authorities, the national competition authorities, the authorities in charge of cybersecurity, and other relevant sectorial authorities in relation to providers of data sharing services[192].

Second, the Commission could also ensure that risks arising from data processing are eliminated or significantly reduced by applying a second screening to the entities who wish to obtain data or insights from the data space, in particular by asking them to submit an application where they explain in detail the way in which they intend to use such data[193]. Depending on the risks involved in the intended use, the Commission could refuse or grant an authorization, which in turn may specify the specific conditions in which access and use of data is permitted. These conditions may include, for instance, the need to obtain consent from data subjects or to offer an opt-out, the use of differential privacy or other de-identification techniques before using such data transferring it to third parties. In the most extreme cases, access could be granted only *in situ*, in other words accommodating queries for such data within the dataset of the original contributor, which has the advantage of preserving the contextual elements that enhance the value of the data in question[194]. A further option would be to grant access only to one or more algorithms operating within the data space in order to derive insights from collective data, without actually ever accessing such data: this can be accomplished thanks to the advancements of secure multiparty computation and homomorphic encryption.[195]

To determine where on this spectrum the proposed collaboration would fit from a competitive standpoint, the sandbox could include presumptions or other trigger points based on the framework outlined by the Horizontal cooperation guidelines, and use as reference framework for the assessment of data collaboratives the distinction made by the ILO between consumer, producer and worker initiatives outlined in section 1 (thus leaving aside the more amorphous concept of multi-stakeholder cooperatives). This would allow to build a common framework for the evaluation of possible harms and efficiencies arising from such ventures, and the identification of further subcategories that facilitate the competitive assessment. Within producers´ collaboratives, for instance, these could include a distinction between: facially pro-competitive initiatives focused on increasing the efficiency of ad intermediation (one example

[192] DGA, supra n. 151, art. 12 (3)

[193] This is similar to what has been proposed by Prufer and Graef (2021), but more focused on the process of data processing declared by data access seekers than the likely outcome of the intended use. See Prufer, Jens and Graef, Inge, Governance of Data Sharing: a Law & Economics Proposal (January 22, 2021). TILEC Discussion Paper No. 2021-001, CentER Discusson Paper No. 2021-004, Available at SSRN: https://ssrn.com/abstract=3774912 or http://dx.doi.org/10.2139/ssrn.3774912

[194] Parker, Geoffrey and Petropoulos, Georgios and Van Alstyne, Marshall W., Platform Mergers and Antitrust (January 10, 2021). Boston University Questrom School of Business Research Paper No. 376351, Available at SSRN: https://ssrn.com/abstract=3763513 or http://dx.doi.org/10.2139/ssrn.3763513

[195] Katrina Liggett and Kobbi Nissim, Ground Rules and Goals for Data Co-ops (2019), paper presented at the Privacy Law Scholars Conference- Europe, University of Amsterdam (2019), on file with the author.

being Adobe Cloud Device Coop[196]), where marketers pool anonymous authentication data and HTTP header information to reach the same customers across channels; initiatives that involve the sharing of consumer characteristics by offering the opportunity to share anonymized data and glean insights from the collective pool of information (two examples being Alliance Gravity[197] and Criteo Commerce Marketing Ecosystem[198]), which could potentially lead to uniform practices across the industry; and more delicate research-focused collaboratives, which can only be admitted upon verification of compliance with key ethical and legal principles, including the limits imposed by data protection law to ensure that prevent the processing of data to trigger "secondary" commercial actions[199]. At the same time, similar principles, and the risk of their violation, could limit the way in which the Commission may wish to encourage competition and efficiency among producers, in particular if consumers do not value the engagement in more targeted advertising as a parameter of competition and differentiation in the market.

Within consumer collaboratives, in turn, one could distinguish between those providing the tools for individualized data sharing (as, for instance, Midata) from those that only enable the sharing on a de-identified or "pseudonymous" basis (as is the case for Citizenme, Meeco, Digime): the former type reveal a greater potential to leveraging by the acquirers into secondary markets, and thus its assessment should be conducted in tandem with an evaluation of the genuineness and granularity of the consent that lies at the basis of the data sharing. Both types of collaboration promise to overcome the "mainstreaming" of content that is associated with ad-funded media (as recognized early on by the founders of Google)[200], and this could be considered as a cognizable efficiency. At the same time, limits could be imposed to the first type of cooperation within data spaces, in order to minimize the risks of misuse of the data, to the determine of data donors but also of related third parties: the more the revealed data can be accumulated over time and used to target consumers on the basis of their characteristics, the more this could lead to steering and manipulation of consumer choice, which could be recognized as a relevant risk for the Commission´s decision to allow or extend the operation of a given collaborative within the sandbox. In other words, the risk of future violations of data protection, cybersecurity and consumer protection law should gravitate against admission to the sandbox, or in favor of the imposition of stringent conditions for admission, and a possible withdrawal as soon as doubts arise about the continued adherence to such conditions. The creation of a data trust, with the appointment of an independent third party for the administration of a data collaborative, would offer guarantees in this respect, especially if accompanied by a process of certification and monitoring approved by the Commission.

Finally, when it comes to workers collaboratives, at least two different initiatives should be distinguished. First, the use of data as a means of intelligence and coordination among workers; and second, the coordination on the disclosure of data as an output of labor. The first is a consequence of the evolution of the concept of "union" that was pointed out in section 1: in order to understand the modern concept of data union, it is important to look beyond the narrow

---

[196] Adobe Launches Experience Cloud Device Co-op (28 March 2018), available at <https://news.adobe.com/news/news-details/2018/Adobe-Launches-Experience-Cloud-Device-Co-op/default.aspx> accessed 01 June 2021

[197] <https://www.alliancegravity.com/> accessed 01 June 2021

[198] <https://www.criteo.com/news/press-releases/2017/10/criteo-empowers-retailers-and-brands-successful-future/> accessed 01 June 2021

[199] See in this sense Giulia Schneider, Health Data Pools under European Policy and Data Protection Law: Research as a New Efficiency Defence? 11 (2020) JIPITEC 49 para 89.

[200] Sergey Brin and Lawrence Page, The Anatomy of a Large-Scale Hypertextual Web Search Engine (1996), available at <http://infolab.stanford.edu/~backrub/google.html> accessed 01 June 2021

focus of institutionalized labor movements, recognizing that data *about labor* is offering workers the ability to coordinate more swiftly and effectively, without necessarily resorting to strikes or other more traditional techniques of workers' representation that are pursued by labor unions. This is a contentious area because coordination may be considered anticompetitive, to the extent that workers fall within the notion of undertakings: this has led the Commission to recently launch a consultation with the aim to ensure that EU competition rules do not stand in the way of collective bargaining by certain solo self-employed people[201]. Based on the information revealed during the consultation, it is likely that the modernized framework will recognize that workers coordination (including exchanging information within the "European skills data space") is beneficial as a countervailing measure to the exercise of market power by management, in particular to prevent the unfair imposition of conditions on a take-it-or-leave-it basis. It could also benefit workers in other ways, by raising collective awareness over the benchmarks used to evaluate their performance and the dynamics of competition in their profession. However, the framework should also recognize limits to this countervailing power, as a reflection of power in the labor market, and by consequence the application of limits with regard to the use of labor data. The second type of worker collaboratives, in turn, ventures into a deeper level of coordination and control over the workers´ output by creating the technical means to withdraw the supply of *data as labor*. When used as a mean of coordination among workers, this involves a coordination that directly affects output (not merely as a result of a negotiation process), and should accordingly be subject to a stricter scrutiny: this will involve an assessment not only of the market conditions and the intended use of the data, but also of the way the collective data is concretely put to use, and the competitive effects that this generates- a complication that might be sufficient to warrant the non-admission of these collaboratives from European data spaces.

In conclusion, it has been shown that innovation hubs and regulatory sandboxes offer a convenient and amenable set of tools for the European Commission to advance the mission of data spaces by enhancing legal certainty and encouraging participation by players that are aligned with the values pursued by this policy innovation, while maintaining close supervision over the products of the selected collaborations. They would allow the Commission to learn about the advantages and pitfalls of various innovative business models and to facilitate the emergence of pro-competitive, secure and trustworthy data collaboratives within European data spaces.

[201] European Commission Press Release: Competition: Commission invites stakeholders to provide comments on the application of EU competition law to collective bargaining agreements for self-employed (5 March 2021), available at <https://ec.europa.eu/commission/presscorner/detail/en/ip_21_988> accessed 01 June 2021.