FROM TIPPING TO TRUSTEES: WHY DATA-DRIVEN MARKETS REQUIRE INSTITUTIONAL DESIGN, NOT OPTIMIZATION

BY JENS PRÜFER & PAUL DE BIJL[1]

1 Tilburg Law and Economics Center and Department of Economics, Tilburg University, the Netherlands, and Centre for Competition Policy and School of Economics, University of East Anglia, UK; j.prufer@uvt.nl. ACM (Netherlands Authority for Consumers and Markets) and Tilburg Law and Economics Center, Tilburg University, the Netherlands. We are grateful to Inge Graef and Martine Postma for very helpful comments. All errors are our own.

CPI ANTITRUST CHRONICLE March 2026

Visit www.competitionpolicyinternational.com for access to these articles and more!

FROM TIPPING TO TRUSTEES: WHY DATA-DRIVEN MARKETS REQUIRE INSTITUTIONAL DESIGN, NOT OPTIMIZATION

By Jens Prüfer & Paul de Bijl

Data-driven markets, where innovation costs decrease with accumulated user information, tip toward durable monopoly even absent predatory conduct. The EU's Digital Markets Act correctly mandates data sharing by gatekeepers but excludes evolving dominant firms and limits obligations to designated platforms. Voluntary alternatives, such as industry-led data spaces, fail structurally: data-rich incumbents earn monopoly rents from data hoarding and face zero incentive to share genuine insights. This article argues that restoring competition requires institutional design, not quantitative optimization. Drawing on economic governance theory, we show that mandatory data sharing demands three institutional features: regulatory adjudication by national competition authorities to determine sharing obligations, enforcement via public data-pooling infrastructure, and judicial review to prevent overreach. Institutions grounded in regulatory legitimacy and coercive enforcement, rather than multi-sided platform pricing mechanisms, are necessary to implement a feasible solution to enforce compliance and restore innovation incentives on data-driven markets.

I. THE INTELLECTUAL FOUNDATION: FOUR PAPERS, ONE NARRATIVE

Over the past fifteen years, a coherent research program has emerged explaining how data-driven competition deviates fundamentally from traditional competition theory. Four papers anchor this narrative.

Argenton & Prüfer (2012)[2] was the first paper in economics about the impact of data on markets: they modeled search engines as subject to *data-driven indirect network effects*. Unlike traditional network effects, where users value others' presence directly, here the mechanism operates through supply-side learning. More users today generate query logs (click data revealing user preferences) that reduce tomorrow's cost of producing quality for all users. Critically, these logs are proprietary (each search engine's success depends on exclusive access to its own user feedback). Constructing a simple model, the authors demonstrated that markets exhibiting this structure tip persistently: even a modest initial lead compounds, as the dominant firm matches rivals' innovations at lower cost (via data advantage), deterring investment by competitors. They also suggested a novel policy proposal: mandatory sharing of anonymized query logs to sever the exclusive link between market share and innovation cost.

Prüfer & Schottmüller (2021)[3] generalized Argenton & Prüfer (2012) beyond search. In their dynamic R&D model, any market where marginal innovation costs decrease in user information and user information is collected automatically, via machines, is *data driven*. Crucially, they introduced the concept of *connected markets*: user information from Market A (e.g., location data from maps) inputs quality in Market B (ride-hailing), explaining Big Tech's conglomerate strategy. They proved that, given sufficiently strong data-driven indirect network effects, monopolization is a uniquely stable equilibrium even without predatory conduct. Smaller firms, foreseeing imitation by data-rich incumbents, rationally exit (or never start business). The authors showed that mandatory data sharing eliminates tipping by equalizing costs. However — and this is vital for our critique below — they assumed a benign regulator implementing sharing in a coordinated, frictionless manner. The question of *institutional design* to achieve this coordination remained untouched.

Klein et al. (2025) provided rare empirical proof. Using the Cliqz search engine and detailed experimental data, they decomposed query traffic into common (high-frequency) and rare (long-tail) queries. For common queries (26 percent of traffic), algorithms converge in quality and query-specific data matters little. However, for rare queries (74 percent of traffic), user-generated data becomes the decisive margin: the measured quality lead of Google over Cliqz among these rare queries amounts to 20 percent, which substantiates many users' perception, that Google is "the best search engine" — but this lead is directly traceable to log-based machine learning, not algorithmic superiority. This experiment vindicates Argenton-Prüfer's mechanism: data hoarding creates measurable consumer harm, not through price or exit, but through quality degradation for users with specialized information needs and market power for the dominant provider.

Finally, Graef & Prüfer (2021) shifted from theory to governance.[4] They argue that neither traditional competition law (limited to abuse of dominance *ex post*) nor GDPR (concerned with individual privacy, not market structure) adequately address data-driven market tipping. They pick up the proposed remedy of mandatory data sharing but recognize it requires careful institutional design. Their governance proposal centers on a European Data Sharing Authority ("EDSA"), a novel public body where firms submit data and algorithms train collectively, with national competition authorities ("NCAs") adjudicating disputes and defining the scope of sharing obligations. This marks a crucial shift: the solution is not a contract, not a market, but an institution backed by law.

II. DATA-DRIVEN NETWORK EFFECTS ACROSS MARKETS

Prüfer and Schottmüller's "connected markets" concept shows why data-driven dynamics extend far beyond search. Data quality feeds innovation across sectors, and Big Tech's conglomerate model incorporates this advantage for exclusive purposes. A problem is that technology companies may be unwilling to share data, in particular when there is a risk that receiving parties may develop into competitive threats.[5]

2 Argenton, Cedric & Jens Prüfer (2012). "Search Engine Competition with Network Externalities," *Journal of Competition Law & Economics*, 8(1): 73-105.

3 Prüfer, Jens & Christoph Schottmüller (2021). "Competing with Big Data," *Journal of Industrial Economics*, 69: 967-1008.

4 Graef, Inge & Jens Prüfer (2021). "Governance of Data Sharing: A Law & Economics Proposal, *Research Policy*, 50: 104330.

5 See Van Gorp, N., De Bijl, P., Graef, I., Molnar, G., Peeters, R., & Regeczi, D. (2020), "Exploring data sharing obligations in the technology sector", report by Ecorys, e-Conomics and Radicand Economics commissioned by Ministry of Economic Affairs and Climate Policy, the Netherlands.

De Cornière & Taylor (2025) provide the general framework.[6] In their model, data's competitive effect depends on whether it exhibits *log-supermodularity*: Does a firm's innovation productivity increase more steeply as it acquires more data?[7] If yes, data-driven markets generate tipping. They also distinguish *revenue-shifting effects* (data enables better targeting, raising prices) from *consumer-surplus effects* (data improves quality). Mandatory sharing can harm welfare if data shifts surplus to competitors without improving service quality. But in quality-sensitive markets (search, e-commerce recommendations, financial advice), data's primary value lies in service improvement, making sharing efficiency-enhancing.

Financial services illustrate this threat. Calef et al. (2024) document how Big Tech's unique datasets, including information on users' browsing history, geolocation, and biometric authentication, create competitive advantages in credit scoring, product targeting, and payment gatekeeping.[8] Apple Pay integrates transaction data with biometric authentication; Google Wallet combines search queries (intent signals) with location data. Traditional banks cannot match this without access to non-financial behavioral data. The FCA acknowledged tipping risk but proposed only monitoring and sandbox testing. This is insufficient given Prüfer/Schottmüller's prediction that, once markets tip, reversal is prohibitively costly.

The lesson is that data-driven dynamics are not a search-engine quirk; they threaten any market where learning (from user behavior) reduces the cost of serving customers. This includes telecommunications (network quality), e-commerce (product recommendations), mobility (route optimization), health (patient insights), and insurance (risk underwriting). The domino effect is real.

III. THE EU'S FIRST STEP: DMA ARTICLE 6(10-11)

The EU's Digital Markets Act, enforced from March 2024, mandates that gatekeepers (firms with >EUR7.5b turnover or 45+ million active end users in the EU) share user data with business-user competitors upon request. Obligations pertain to data access for business users to their own data, end-user data portability, search engine data sharing with third-party search engines, and advertising performance data. This was a necessary first step in response to urgent problems, but is structurally insufficient as there are (at least) three critical gaps (see Prüfer, 2025):[9]

1. The DMA's quantitative gatekeeper thresholds exclude a range of mid-tier firms that may already display data-driven tipping dynamics, for example firms with around 30–50 percent market share in a data-driven market. These firms are not captured by the gatekeeper definition, even though from a competition-economics perspective they may already enjoy a data-driven entrenched position. This calls for lowering effective thresholds and treating mid-tier dominators differently from truly small players. This logic is consistent with the broader simplification agenda behind the Digital Omnibus, which carves out SMEs and "small mid-cap" firms from some digital obligations to boost competitiveness and suggests more differentiated treatment of smaller players.[10]
2. DMA Article 6(10-11) applies only to designated gatekeepers. But data-driven dynamics operate across supply chains. A firm may not meet gatekeeper criteria yet still hoard crucial user insights that erect barriers for competitors. Cross-market mandatory sharing may better address the connected-markets problem. E.g., financial institutions with >30 percent market share should share customer behavioral data with authorized fintech firms.
3. The GDPR's consent framework requires data subjects' explicit authorization for many secondary uses of personal data. DMA Article 6(9)–(10), in turn, imposes a legal obligation on gatekeepers to make certain data accessible, but it does not displace or weaken GDPR protections. In practice, this interaction can still create consent bottlenecks for data recipients and for cross-sector sharing beyond the DMA's narrow scope. Clarifying, within the GDPR framework, that competition-mandated data sharing may in appropriate cases rely on legal bases such as Article 6(1)(e) (tasks carried out in the public interest) or 6(1)(f) (legitimate interests), while fully respecting purpose limitation and data-minimisation principles and focusing on anonymized or non-sensitive data, would reduce frictions without lowering the level of privacy protection. These structural gaps demonstrate why mandatory sharing (as the DMA correctly recognizes) is necessary but must be broadened: voluntary data-sharing initiatives, such as industry-led data spaces, cannot be expected to overcome

6 de Cornière, Alexandre & Greg Taylor (2025). "Data and Competition: A Simple Framework," *RAND Journal of Economics, 56(4):* 494-510.

7 The advantage stems from mix of economies of scale (amount of data), economies of scope (variety in applications of data) and learning effects (more data generates more knowledge).

8 Calef, A., S. Ennis, B. Enstone, & J. Prüfer (2024). "Consultation Response to FCA Call for Input: The Potential Competition Impacts of Big Tech Entry and Expansion in Retail Financial Services," Centre for Competition Policy, University of East Anglia, January 2024.

9 Prüfer, Jens (2025). "Consultation Response: Joint EC/EDPB Guidelines on DMA-GDPR Interplay," Tilburg Law and Economics Center and Centre for Competition Policy, University of East Anglia, December 2025.

10 https://www.europarl.europa.eu/news/en/press-room/20260220IPR35906/simplified-rules-for-small-mid-cap-companies

dominant firm's rational incentives to hoard data, while the DMA's gatekeeper-only scope leaves mid-tier dominators (30–50 percent share) unconstrained. See Meyers (2026) for a broad exploration of ways to effectively implement the DMA's rules and obligations, concluding that no single governance model can tackle all DMA obligations or all types of platform services.[11] In what follows, this paper focuses on the institutional context to further enforce and stimulate data sharing.

IV. THE LIMITS OF VOLUNTARY DATA SPACES ON DATA-DRIVEN MARKETS

An alternative and complement to mandatory data sharing as currently imposed by the DMA is voluntary sharing via data spaces, quite close to the proposed data pooling in Graef & Prüfer (2021) and to the private-sharing model envisaged by the Data Governance Act ("DGA"). The purpose would be to complement the arrangements currently included in the DMA, in order to stimulate innovation through increasing the possibilities for using and combining data. . In a legislative package called the 'Digital Omnibus', the European Commission aims at streamlining various 'digital acts', including the GDPR, Data Act ("DA") and DGA.[12] Under this package, the DGA will get a more voluntary nature. The DMA is not part of this package but may be subjected to a structural review in the near future. Note that the DA, unlike the DMA, explicitly addresses data from smart devices (hardware) and integrated software. Designated gatekeepers under the DMA can also be manufacturers of smart devices, but the type of data tends to be different.

Data spaces would complement the arrangements currently included in the DMA, in order to stimulate innovation through increasing the possibilities for using and combining data. Here, several data controllers bring their data sets together in order to learn from the combined sets and benefit from data economies of scale and scope (Carballa-Smichowski et al., 2025).[13] How can one make this work? Bisière et al.'s (2025) policy paper on data spaces sketches an idea for a solution.[14] These authors view data spaces as multi-sided platforms with decentralized governance, where cost-sharing mechanisms (implemented via Shapley-value allocations – see below) balance participation incentives. At first sight, the paper's framing looks appealing: data spaces are described as evolving institutions for voluntary cooperation, avoiding regulatory burden. In specific settings, such a solution might work. The question is whether it can work in a dynamic setting with a high rate of technological change and a large number and variety of participants, which add complexity and uncertainty to the mix. By contrast, emerging sectoral 'regulatory' data spaces, such as the European Health Data Space, already rely on binding access obligations and specialized data-access bodies, which fits the institutional solution argued for in this paper.

Bisière et al.'s characterization of data spaces interestingly shares similarities with the characterization of *cooperatives* (of producers, workers, consumers, or capital-owners) in Henry Hansmann's (1996) *magnum opus*.[15] There, Hansmann distinguished two organizational forms:

Cooperatives, where contracting costs are minimized if patron-owner members (equal in principle, with votes/profits proportional to patronage) are relatively homogeneous and all are transaction-intensive. Dairy farmers' cheese cooperative exemplifies this: all members have comparable scale, all contribute regularly, all benefit proportionally. The sharing mechanism can effectively handle uncertainty (e.g. due to fluctuations in milk volumes) and will hardly have to deal with unanticipated consequences of new technology.

Investor-owned firms (equal to capital cooperatives in Hansmann's framework), where capital contributors control the organization via residual claim rights. This organizational form minimizes contracting costs when patron heterogeneity requires centralized coordination. As an example, a large supermarket chain's supply chain involves vastly different-scaled suppliers. Hence, central coordination minimizes total costs, rendering investor-owned firms efficient. Investment mechanisms are designed to deal with the uncertainty from failure or success by firms, which can simply be measured by quantitative measures such as cash flow and profits.

Bisière et al. (2025) define data spaces as cooperative-inspired, which.h can be viewed as a perspective in the spirit of Hansman (1996): members retain data sovereignty, collaborate on services, and share value equitably. Their analytical toolkit is the multi-sided platform

11 Meyers, Z. (2026). "Open Tech Platforms: Technology and Governance Mechanisms", Issue Paper, CERRE.

12 Digital Omnibus Regulation Proposal | Shaping Europe's digital future.

13 Carballa-Smichowski, Bruno, Néstor Duch-Brown, Seyit Höcük, Pradeep Kumar, Bertin Martens, Joris Mulder & Patricia Prüfer (2025). "Economies of scope in data aggregation: Evidence from health data,"Information Economics and Policy, 71: 101146.

14 Bisière, Christophe, Jacques Crémer, Bruno Jullien & Yassine Lefouili (2025). "The Economics of Data Spaces," TSE Policy Paper, July 2025.

15 Hansmann, Henry (1996). The Ownership of Enterprise, Cambridge: Harvard University Press.

(MSP) model (Rochet & Tirole, 2006),[16] which is often used to analyze digital platforms such as Uber (drivers/riders), Airbnb (hosts/guests), and payment systems (merchants/consumers), amongst others. In MSP theory, an orchestrator implements a governance system and then balances asymmetric pricing, for instance by subsidizing high-externality sides to attract them, to maximize aggregate value creation by platform users (and then skim off the cream without discouraging valuable user interactions).

Note that the governance system of a platform does not constitute cooperation, even though it may be designed to facilitate (decentralized) cooperation by its users, such as in a sharing platform. The design of a platform is an expression of hierarchical coordination by a central player (the platform owner). By contrast, a cooperative would operate via democratic governance, e.g., assigning one-member-one-vote rights, or votes weighted by members' relative contributions. An MSP operates via unilateral pricing authority by the orchestrator. These are different organizational forms.[17]

Moreover, Bisière et al. (2025) propose allocating costs using Shapley value, which distributes total value based on each member's marginal contribution to all coalitions. The underlying idea is to circumvent risks of free riding on each other's contributions. To that end, the Shapley value is "fair", in the sense that (by the definition of the concept) each member receives a payment equal to their average marginal contribution. But Shapley value computation requires two inputs that are hard to achieve or obtain in practice: complete information about all possible coalitions' outputs, and an impartial auditor computing values and enforcing allocations. First, in data spaces, this means: if firms A, B, and C pool data, how much joint value does their data create – now and in the future? What if A and B are alone, etc.? In practice, data's value is emergent and recipient-specific, which makes it easy for firms to over-estimate their own data's value and to under-estimate their partners'. Disputes are very likely, even more so in the light of data applications that have yet to be developed, while participants may have strong thoughts about them.

Second, who could serve as an impartial auditor? If the data space orchestrator computes Shapley values, it has an incentive to bias allocations in its favor. If an independent auditor computes them, governance costs increase steeply, and the flexibility advantage of data spaces (vs. traditional firms) evaporates. Such problems will not disappear if one sets the platform apart from its users, for instance as a separate not-for-profit entity. The owners (possibly the intended users) will face the same challenges. Note that the paper points out that in B2B settings, the number of participants will typically be relatively small, which would facilitate implementation of the proposed solution. However, to generate maximal value from data, one wants to invite many parties on board, including - in particular - start-ups. Thus, there is an inherent tension between the potential of the positive externalities of data sharing and the feasibility of the mechanism.

In practice, the Shapley value cannot avoid reflecting the underlying power asymmetry. The orchestrator is likely to invoke "fairness" while pricing in a way that maximizes its own rent (or the rents of owners with the largest vested interests). Building a second governance layer, to "audit the auditor," may reduce the orchestrator's bias but drive up governance costs even more, which begs the question whether this route will lead to an optimal design.

Therefore, even if governance were perfectly designed, voluntary data sharing via data spaces may fail on data-driven markets due to incumbent incentives and inherent conflicts of interests.

This is what Graef & Prüfer (2021) already addressed: dominant data-rich firms have no incentive to share because they already profit tremendously from their data advantage. A firm with, say, 60 percent market share earning quasi-monopoly returns faces a choice. Either share data voluntarily, which implies losing competitive advantage, watching rivals improve quality, and facing declining margins. Or to hoard data without sharing, maintaining advantage, deterring entry, and keep extracting monopoly rents.

Rational economic actors will be very hesitant to share data or drastically reduce the amount of data they are willing to share – which undermines the whole idea of data sharing. Voluntary data spaces may collapse because the only willing participants are data-poor firms seeking others' data, a configuration that cannot sustain cooperation absent external enforcement. Incumbent firms will join data spaces (for real, not nominally) only if forced by regulation. Even then, they can be expected to minimize data quality shared (submitting degraded, anonymized, or delayed data) unless legal enforcement mechanisms compel compliance.

16 See Rochet, J.-C. & Tirole, J. (2006). "Two-sided markets: a progress report," *RAND Journal of Economics*, 37(3): 645-667.

17 Of course one can organize a platform as a cooperative, with users as platform owners (Wegner et al., 2023). See Wegner, D., da Silveira, A.B., Marconatto, D. and Mitrega, M. (2024). "A systematic review of collaborative digital platforms: structuring the domain and research agenda," *Review of Managerial Science,* 18: 2663–2695. Our impression is that it is hard to make them successful, as platform cooperatives face an inherent tension: the features that make cooperatives attractive (such as democratic participation, member control of surplus distribution) may undermine rapid scaling, decisive investment decisions, and network effects build-up (see e.g. Bunders et al., 2022). See Bunders, D.J., Arets, M., Frenken, K., & De Moor, T. (2022). "The feasibility of platform cooperatives in the gig economy," *Journal of Co-operative Organization and Management*, 10(1): 100167.

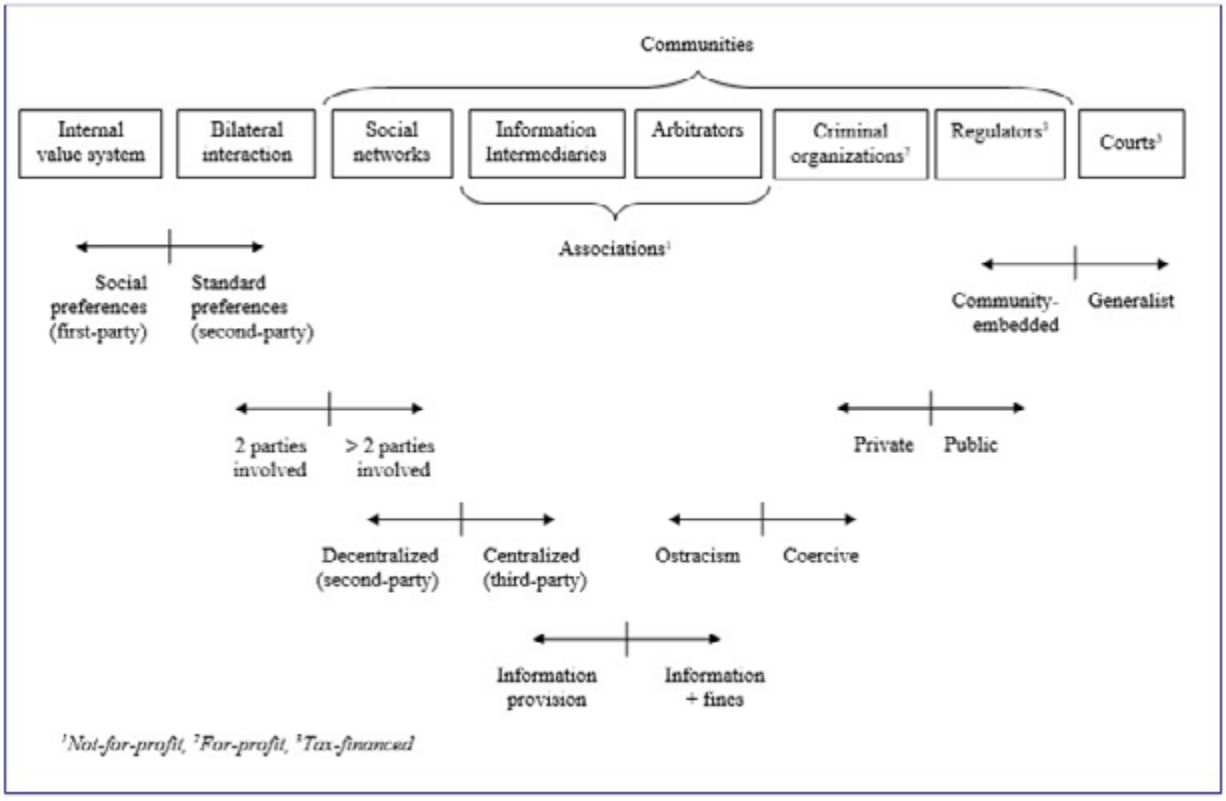
V. AN INSTITUTIONAL SOLUTION: ECONOMIC GOVERNANCE

Prüfer's (2024) "Economic Governance and Institutional Design" provides a useful framework to address the issues discussed above (this part of the analysis has overlaps with Graef and Prüfer, 2021).[18] Rather than viewing data sharing as a contracting problem (negotiable between firms) or a multi-sided platform problem (based on a governance system that includes an 'optimal' pricing structure), it identifies data sharing as an enforcement problem requiring institutional design. The justification for not leaving the solution up to firms lies in the presence of external effects, such as benefits and spillovers from data sharing.

Prüfer's typology of contract enforcement institutions orders eight archetypes along three binary dimensions:

1. Enforcement Technology: Ostracism (social sanctions, hurt over time) vs Coercion (state punishment, hurts immediately)
2. Information Structure: Bilateral/Multilateral (parties adjudicate themselves, observability of actions sufficient) vs Third-Party (external specialist adjudicates and potentially enforces)
3. Scope: Private Ordering (non-state) vs Public Ordering (state-backed)

See the following figure (or the much more comprehensive description in Prüfer, 2024) for details.



Source: Prüfer (2024)

Each institution's efficiency depends on the specific economic governance problem. There is no silver bullet efficiently solving all economic governance problems. For data sharing on data-driven markets, the problem is severe because of three main reasons.

First, complexity. Determining whether a firm "truly shared" high-quality data vs. degraded data requires specialized technical and economic expertise. Bilateral negotiations can easily fail. Parties cannot objectively adjudicate among themselves. Second, incentive misalignment. Data-rich incumbents prefer non-compliance. Private-ordering sanctions (ostracism, arbitration) rely on the threat of exclusion. But if a data-hoarding monopolist earns EUR10 billion annually in monopoly rents, the threat of exclusion from a data-sharing collective that is worth EUR1 billion in potential competitive gains is not credible. Third, impersonal exchange. Digital markets lack the close-knit communities where social norms enforce cooperation. Participants interact via algorithms, formal agreements and APIs, not personal relationships.

18 Prüfer, Jens (2024). "Economic Governance and Institutional Design," TILEC Discussion Paper 2024-10, Tilburg University.

All three of these features push us to the right of the typology: public ordering, coercive enforcement, third-party adjudication. Among public-ordering institutions, regarding the enforcement of mandatory data sharing, *regulators* are superior to generalist courts. First, regulators employ economists, data scientists, and lawyers who understand data-driven markets. Courts lack this specialization. Second, regulators can craft industry-specific mandates (data-sharing obligations for search differ from those for financial services). Courts apply general legal principles mechanistically. Third, court proceedings can take years, but markets move in months. Regulators can issue guidance and adapt rules more rapidly.

The downside of regulators and the rules they have to enforce is the risk of capture by entrenched interests (including those of Big Tech firms). Industry incumbents lobby for favorable rules. Prüfer & Graef (2021) address this threat via organizational governance, by separating adjudication of a specific data-sharing case (determining if data sharing is mandated) from its enforcement (monitoring compliance) and by building in various checks and balances. Specifically, they propose that national competition authorities ("NCAs") investigate whether a market is data-driven and define sharing obligations. NCAs are staffed by competition lawyers and economists, elected officials appoint leadership, and courts review decisions for legality. This separation from enforcement reduces capture (but does not shield politicians and policy makers from undue influence on the design of the rules to be enforced).

Complementarily, enforcement of data-sharing obligations should be sourced out to a Data-Sharing Authority: a public agency at the EU level or sector-specific regulators nationally, which operate (or supervise) a collective data-pooling infrastructure. Large firms submit data (in case of personal data: anonymized, de-identified per GDPR) and many small firms can send their machine-learning algorithms into the pool and receive access to their algorithm's insights from the pooled data.[19] [20] The Authority monitors compliance through technical audits and algorithmic verification (e.g. checking if submitted data's quality matches historical submissions). If a firm contests an NCA finding or an action by the Data-Sharing Authority, general courts review the decision for legality and proportionality and can annul or adjust it where necessary. This judicial control protects firms against regulatory overreach while preserving effective enforcement.

Bisiére et al. (2025) view governance as an optimization problem: given firms' data and the structure of the market, compute a Shapley value that incentivizes participation. This framing boils down the governance problem into a technical exercise.

By contrast, Prüfer's (2024) framework inverts the question: What institutional structure makes cooperation self-enforcing? The answer emphasizes hard-to-quantify elements, such as trust, incomplete contracts, and legitimacy. Can firms actually trust that competitors receiving shared data will not exploit it (e.g., learning trade secrets embedded in behavioral data)? Moreover, data-sharing agreements must accommodate unanticipated future uses. A rule that today seems fair may become unjust if markets evolve. Institutional flexibility (regulators revising rules) is preferable to contractual rigidity. Finally, will firms comply voluntarily (or at least without gaming compliance) if they perceive an adjudicator or enforcer as legitimate? Public regulators, accountable to elected officials, enjoy greater legitimacy than private orchestrators.

VI. POLICY RECOMMENDATIONS

Based on this analysis, several directions for possible reform of today's data-sharing frameworks come to mind:

1. Determine data-drivenness per market and align the DMA with the DA: On data-driven markets, authorities should first test whether innovation quality and cost materially depend on user-generated data (e.g. along the lines of Klein et al., 2025).[21] Where this holds, the DMA can address structurally powerful gatekeepers, while the Data Act empowers users of connected products and related services to obtain and share their device and service data with third parties of their choice. This user-controlled access model is not sufficient to solve the structural tipping problems analyzed in this paper, but it can complement mandatory sharing by lowering frictions for smaller rivals.

2. Clarify GDPR-DMA Interplay while reaffirming data sovereignty: Competition-mandated data sharing must stay fully within the GDPR. Guidance and horizontal instruments such as the Digital Omnibus should spell out safeguards for processing when firms com-

19 By 'submitting data' we mean, more generally, providing access to data, for instance through APIs allowing access seekers real-time access to the same data as used by the provider, and real-time continuous updating of provided data.

20 Relevant data may encompass user data, user-generated data, machine-generated data etc. Different types of rules and legislation may apply to different data types, ranging from privacy rules to intellectual property rights.

21 Klein, Tobias, Madina Kurmangaliyeva, Jens Prüfer & Patricia Prüfer (2025). "How Important Are User-Generated Data for Search-Result Quality? Experimental Evidence," *Journal of Law & Economics*, 68(3): 499-518.

ply with data-access obligations. Individuals' right not to share should be explicit, particularly for personal data; anonymized or truly non-personal data can circulate more freely, subject to purpose limitation, data minimization, and strong protections against re-identification and unwanted secondary use.

3. Establish Data-Sharing Authorities building on the DGA: The DGA already facilitates trustworthy *private* data sharing via intermediaries (possibly in the form of data spaces). *Public* Data-Sharing Authorities should complement these DGA schemes where access is mandatory (as in Graef/Prüfer, 2021), managing collective pooling with clear mandates and transparent governance, so that voluntary DGA arrangements and competition remedies work in tandem rather than at cross-purposes.

4. Impact Monitoring: Mandate periodic studies (à la Klein et al. 2025) examining whether shared data measurably improves service quality for competing firms. If sharing remains cosmetic, escalate to higher enforcement levels.

5. International Coordination: Work with UK, Canada, and other jurisdictions to harmonize data-sharing mandates, preventing jurisdictional arbitrage where Big Tech firms route data through compliant regions. Realize that incentives of countries where Big Tech firms reside may be different.

Notably, these proposals speak directly to several comments made by industry stakeholders in response to the European Commission's review consultation of the Digital Markets Act (responses published January 2026), which overwhelmingly identified three priorities.[22]

First, respondents called for clearer DMA Article 6(11) guidance and extension to AI training datasets, model parameters, and cloud-stored data. Many emphasized that proprietary datasets controlled by gatekeepers constitute the decisive barrier to competitive AI development. This is precisely the data-driven tipping mechanism documented in Section I.

Second, respondents noted that quantitative thresholds fail to capture hyperscaler dominance (AWS, Azure, Google Cloud) and that AI/LLM services either fall within existing Core Platform Services or require new categories. Lock-in via cloud dependencies and deep integration of gatekeepers' own generative AI models exemplify the connected-markets problem analyzed in Section II.

Third, complaints focused on slow procedures, circumvention tactics, and user-interface designs that undermine choice. Proposed remedies include binding timelines, interim measures, independent audits, and structural remedies for systemic non-compliance. These are institutional features consistent with the regulatory framework advocated in Section V.

Tellingly, gatekeepers themselves resisted expansion of the DMA, arguing that current obligations are already costly and inflexible. This divergence between gatekeepers advocating "dialogue-based enforcement" (6 out of 7 replied) and other stakeholders (300+ from SMEs, business users, civil society, and academics) demanding stronger mandatory data sharing, confirms the structural incentive misalignment identified in Section IV: firms earning monopoly rents from data hoarding will rationally oppose any regime that equalizes access.

VII. CONCLUSION

This article synthesizes fifteen years of research on data-driven competition and adds an institutional perspective. Four foundational papers established the problem: markets where innovation costs decline with user information tip persistently toward monopoly. Even modest initial leads compound, as dominant firms match rivals' innovations at lower cost via proprietary data advantage. Empirical evidence confirms that 70+ percent of search traffic exhibits this dynamic, with quality gaps traceable directly to data hoarding rather than algorithmic superiority. Traditional competition law and privacy regulation prove insufficient; neither addresses the structural incentive for incumbents to hoard and not share data.

The EU's Digital Markets Act represents progress by mandating that gatekeepers share data. But three gaps limit its current effectiveness: the gatekeeper definition excludes evolving dominant firms; obligations apply only to designated platforms, not across supply chains, whereas GDPR consent requirements create bottlenecks for data recipients. Moreover, voluntary data-sharing initiatives cannot succeed. Industry-led data spaces, even when governed via sophisticated cost-allocation mechanisms, face an insurmountable obstacle: firms earning billions from data monopolies will rationally refuse to cooperate in sharing, as any participation erodes their competitive advantage.

22 https://digital-markets-act.ec.europa.eu/consultation-first-review-digital-markets-act_en

The solution lies in institutional design. Economic governance theory identifies three conditions under which mandatory sharing becomes self-enforcing. First, complexity and incentive misalignment push enforcement from private ordering (social norms, arbitration) toward public ordering with coercive enforcement. Data-sharing disputes require specialized technical and economic expertise that private arbitrators lack, and incumbents' monopoly rents render ostracism threats non-credible. Second, separating adjudication from enforcement reduces capture risk. National competition authorities should determine when markets are data-driven and define sharing obligations; a public Data-Sharing Authority should operate collective data-pooling infrastructure and monitor compliance. Third, judicial review ensures regulatory actions remain proportionate and legally sound.

This institutional architecture supports trustworthy relationships, incomplete contracting, and legitimacy over optimization under incomplete information and unclear goals of the players involved. Firms comply not because a Shapley value appears "fair" but because a legitimate public authority credibly commits to enforcement. Regulators, accountable to elected officials and subject to judicial review, have legitimacy that private orchestrators cannot replicate. Incomplete contracts accommodate unanticipated future uses of shared data; institutional flexibility allows regulators to revise rules as markets evolve, avoiding contractual rigidity.

The path forward requires recognizing that data-driven competition is fundamentally an institutional problem. Markets are fragile and prone to monopolization when innovation depends on proprietary information. Voluntary coordination will tend to fail because incentives misalign. The solution is not more complex contracts, better algorithms or fairer pricing. It is building public institutions with the expertise, legitimacy, and coercive power to enforce sharing obligations. Only such institutions can restore innovation incentives and stimulate decentralized and parallel innovation processes on data-driven markets.

CPI COMPETITION POLICY® INTERNATIONAL

CPI Subscriptions

CPI reaches more than 35,000 readers in over 150 countries every day. Our online library houses over 23,000 papers, articles and interviews.

Visit competitionpolicyinternational.com today to see our available plans and join CPI's global community of antitrust experts.