Intellectual Property Theories as Applied to Big Data

Mamanazarov Sardor Shukhratovich
Tashkent State University of Law
sardormamanazarov@gmail.com
ORCID: 0009-0004-5855-6498

Abstract

Emerging data-driven technologies and exponential information growth are catalyzing new conceptualizations of data governance. This article undertakes comparative analysis of intellectual property regimes and rights theories as applied to big data. It evaluates tensions between public good knowledge sharing and private interests in data control. Doctrinal, ethical, economic and policy perspectives inform examination of varied models including IP analogues, effort-based rights, unfair competition, metadata protections, technological controls and context-specific data rights. The study assesses merits and critiques of existing and proposed regulatory approaches across jurisdictions. It concludes by proposing balanced frameworks recognizing collective oversight, differentiated data protections and public interest mandates as pathways to stimulate socially beneficial data innovation.

Keywords: Big Data, Data Governance, Intellectual Property, Open Data, Data Rights, Data Ethics, Data Trusts, Data Regulations

I. Introduction

In recent decades, data has rapidly emerged as a new form of economic asset, presenting novel challenges for intellectual property law frameworks predicated on incentivizing innovation through defined rights over intangible creations. As digital technology enables the exponential proliferation of information that can be collected, analyzed and monetized, devising appropriate policies to govern data access and control has become a pressing priority across various sectors. Conflicting perspectives pit notions of data as a public good supporting knowledge advance against its treatment as proprietary commodities. Resolving these tensions requires re-examining the theoretical underpinnings of intellectual property in relation to the sui generis nature of data [1].

At its core, intellectual property law reflects a carefully calibrated social contract granting creators limited monopoly rights over intangible works in order to stimulate continued innovation and distribute the fruits of human ingenuity. The incentive structures of patent, copyright, trademark and trade secret regimes aim to balance public interests in access with private interests in exclusivity over novel inventions, original expressions, distinctive brands and confidential information respectively. However, applying these conventional IP frameworks to the data
context poses several conceptual challenges. Data lacks the intentional creativity underpinning copyright law [2].

Factual information is not invented but discovered, raising questions about whether data generation involves human agency qualifying for patent-like protections. Data functioning in technical systems differs from trademark's emphasis on branding and consumer identification [3]. The diffuse nature of data flows conflicts with constructs of secrets enabling competitive advantage. Furthermore, the relational and cumulative character of data clashes with IP's focus on discrete finished products. Granting proprietary rights may impede downstream uses that spur innovation. Alternate theoretical bases rooting data value in invested labor, compromised secrets or unfair free-riding have informed emerging sui generis regimes that attempt to conceptualize data control rights [4].

As data permeates every sector of the global economy, devising appropriate governance frameworks balancing public good knowledge exchange with private interests in extracting value from data has become an urgent policy priority. Concerns over fragmented regulatory approaches risk jeopardizing data-driven innovation through either under-protection disincentivizing investments, or over-protection restricting access. This article undertakes a comparative analysis of intellectual property theories as applied to conceptualizing rights over data, evaluates tensions between private monopolies and public goods, assesses merits of existing and proposed regulatory regimes, and articulates principles for developing balanced sui generis data governance frameworks tailored to sectorial needs in furthance of innovation [5].

II. Methodology

This article employs a multidisciplinary doctrinal methodology combining legal analysis of established and emerging data regulations across jurisdictions with conceptual perspectives from information ethics, innovation economics and technology policy. To contextualize the issues, the analysis begins by tracing the contours of conventional intellectual property law frameworks underpinning the patent, copyright, and trademark and trade secret regimes. Seminal statutes, landmark judicial decisions and academic scholarship are reviewed to elucidate the subject matter, requirements, rights and limitations defining current IP protections. The conceptual underpinnings emphasizing incentivizing innovation through delineated exclusivity over products of human creativity are examined to highlight the centrality of originality and secrecy constructs [6].

The analysis proceeds to examine fundamental differences between the subject matter protected under conventional IP laws and the sui generis nature of data posing challenges to analogized application of existing frameworks. The absence of original human creativity in factual data is contrasted with copyright law's emphasis on expression over underlying ideas. The non-rival and non-
excludable qualities of data even when appropriated are distinguished from the secrecy requirements of trade secrets. The lack of novelty in continuously updated dynamic data is differentiated from the invention criteria for patentability. And the functional non-identifying role of data is compared with trademark law's branding focus [7].

To assess the merits of existing and proposed models for conceptualizing rights over data, the analysis categorizes and comparatively evaluates five emergent approaches evident in recent regulatory developments [8]:

- **IP-analogous frameworks asserting rights over data based on invested labor and effort.**
- **Unfair competition models targeting misappropriation and free-riding.**
- **Regimes recognizing metadata rights attributing data originators.**
- **Technology-driven approaches utilizing tools like watermarking and blockchain.**
- **Sui generis national security and sector-specific data rights.**

Under the effort-based rights framework, regulations like the European Database Directive predicating protection on substantial investment independent of originality and creativity are analyzed. Misappropriation focused models like the proposed EU Data Act creating obligations on large platforms are reviewed. Emergent attribution-focused metadata rights like the proposed intellectual property protection for data are discussed. Technical implementations such as digital watermarking and distributed ledger permissions systems are assessed. And national security centric data localization requirements as well as sector-specific data rights regimes in domains like health, IoT and mobility are examined [9].

Each model is evaluated on criteria such as its conceptual coherence, subject matter definition, disclosure and access provisions, remedy frameworks, innovation impacts, duration and limitations. The analysis relies extensively on primary legal sources including legislation, case law, regulatory proposals, government reports, and impact assessments across multiple jurisdictions for an international perspective. Secondary analysis also utilizes scholarly articles, think tank and industry association reports surveying the legal landscape. Beyond conceptualizing data rights, the analysis also critically examines countervailing considerations grounding data policy in public good knowledge sharing frameworks [10].

Theories of information ethics and knowledge commons emphasizing equitable access to non-personal data critical for research, innovation and public welfare are discussed. Drawing on economic and philosophical scholarship, the risks of over-propertization from unbounded IP style private monopolies over data are highlighted along with alternatives like liability rules and compulsory
licensing. To articulate balanced governance recommendations, the analysis evaluates hybrid collective data stewardship regimes that distribute oversight and bargaining power across stakeholders through mechanisms like data trusts and pool-based structures. Proposals for differentiated treatment tying data access rights and exceptions to defined public interest uses are assessed [11].

The analysis relies extensively on emerging multi-disciplinary scholarship at the intersection of law, technology and public policy to develop nuanced perspectives on data regulation supporting innovation and the common good. The doctrinal and conceptual analysis is further enriched through comparative study of data regulations and rights frameworks across key developed economies including the United States, European Union, United Kingdom, Germany, France, Japan, Canada, Australia and Singapore. Reviewing convergence and divergence in adopted and proposed policy approaches provides useful case studies and experiments to identify best practices on issues like scope, compliance burdens, remedy structures and public interest safeguards [12].

Based on the multi-faceted legal and conceptual analysis, the study synthesizes salient tensions between private control and public access at the heart of regulating data innovation. It concludes by proposing balanced policy recommendations and identifying critical open research questions for this emerging area at the cutting edge of law, technology and the knowledge economy. With data permeating innovation across sectors, intellectual property theories reconceptualized for the data context can play a vital role in shaping national policies and international harmonization efforts to equitably govern digital intelligence as a global public good while sustaining incentives for its socially beneficial development [13].

III. Results

A. Limits of Conventional IP Models for Data Rights

Applying the predominant intellectual property regimes centered on copyright, patents, trademarks and trade secrets to conceptualize exclusive rights over data reveals several conceptual limitations and mismatches. Core aspects of data creation and use differ fundamentally from the subject matter, requirements and rights constructs underpinning these conventional IP frameworks. Copyright law protects original creative works of authorship like books, articles, music, films, software and artworks. But data lacks the intentional expressive choices underpinning copyrightable creativity. Factual information is discovered rather than created [14].

While selection and arrangement decisions in curating data involve human judgment, copyright still requires causal agency in rendering the protected work. It shields expressions, not underlying facts and ideas. The seminal US Supreme Court decision in Feist v. Rural Telephone established that pure factual
compilations lack originality qualifying for copyright, absent features like selective presentation reflecting creative choices. The EU Database Directive does enable sui generis protection for database contents based on invested labor and resources rather than creativity, but still predates rights on authorship agency [15].

Further, copyright envisages discreet finished products emerging from creative work. But data exhibits a relational interconnectedness as inputs feeding downstream usage and analysis. Datasets are constantly updated rather than comprising fixed creations. And copyright law prizes public access over protection, given non-rival nature of expressions that can be widely shared even when appropriated. This inclination contrasts with proprietary data access restrictions. Copyright's fair use provisions permitting limited unlicensed usage for research and educational purposes also clash with constrained data rights [16].

Obtaining patents requires demonstrating novel non-obvious inventions with clear industrial utility. But data is inherently about recording factual information rather than inventing it. Insights derived from analyzing data trends represent discovery rather than invention. Patent law emphasizes secrecy enabling competitive advantage over inventions disclosed upon patenting. Much data circulating openly would fail secrecy criteria even if deemed invented. Subsequent data uses generate continuous updates rather than fulfilling patent law's discrete invention requirements. And patent terms are fixed, while data retains value beyond arbitrary protection cut-offs. Patents necessitate specifying inventions in claims functioning as metes and bounds over rights. But data's fluid evolving character eludes such boundaries [17].

Further, patent principles entail disclosing full technical details for replication by experts. But data's value lies in comprehensive aggregates and network effects, not singular discrete nuggets that can be segmented out and disclosed. Trademarks denote source identification in commerce. They represent brands distinguishing products and services rather than signifying the underlying tangible goods themselves. Data's primary function is not branding or indication of commercial origin. Trademarks envisage customer reliance for purchasing decisions. But data serves computational and analytical objectives unrelated to consumer identification of data producers. Obtaining trademark protection requires demonstrating bona fide use in commerce. Much data circulation remains non-commercial in research and other contexts [18].

Renewals necessitate continued use showing evolving brand identity and reputation. But data integrity depends on preventing uncontrolled mutations. Doctrines against trademark genericide from over-broad usage have little applicability to data sharing needs. Trade secrets law protects valuable confidential information affording competitive advantage. But data's intrinsic value lies in circulation rather than secrecy. Competitive value derives from scale, network effects and analytical insights from combining datasets. Public and private actors
collect exchange and publish vast data that clearly falls outside trade secrecy protection [19].

Further, trade secret principles entitle independent creation and reverse engineering. But data's fact-based nature makes independent sourcing identical datasets virtually impossible. And data utility depends on use rather than secrecy. Redundancy improves integrity. Restricted access creates anti-competitive effects. Trade secrets envisage commercial actors. But public agencies extensively generate and rely on data. Doctrines also require reasonable efforts to preserve secrecy. But open public data by definition negates such efforts. Thus, existing intellectual property regimes are intrinsically ill-suited to accommodate data's myriad peculiarities challenging the concepts and requirements underpinning copyright, patent, trademark and trade secret protections [20].

Nonetheless, amidst data's rising economic significance, sole-source control and commodification imperatives have spurred attempts to squeeze data into IP frameworks regardless of the dissonance. This risks impacting innovation through misapplication of IP principles evolved for other contexts. It necessitates developing tailored sui generis data governance frameworks that blend public access and private incentives absent in existing regimes. Various theories have emerged attempting to construct alternative bases for data rights independent of conventional IP systems, even while exhibiting similar proprietary tendencies. The European Database Directive exemplifies a prominent effort-based rights theory rooted in protecting invested labor and resources rather than creativity expressions [21].

Codifying the "sweat of the brow" doctrine, it grants 15 years of protection against data extraction and reuse solely based on demonstrating "substantial investment" in obtaining, verifying or presenting database contents. This subsumes creativity ideals like originality under a broad mantle of commercial data rights, justifying proprietorship through strenuous compilation efforts. However, critics argue that encouraging industry self-interest in hoarding data may impede socially beneficial uses and stifle down-stream innovation that relies on combining data from multiple sources [22]. Factual monopolies with no originality test conflict with the quid-pro-quo balance in IP bargain theories underlying copyright and patents. Challengers note that data collection and aggregation costs continue declining rapidly. Granting rights over indiscriminate database contents rather than novel creations upend public domain ideals. It commodifies raw facts And it disguises free-riding arguments by deeming re-use of even public domain [23].

**B. Emergent Sui Generis Regimes for Data Rights and Control**

Given limitations constraining analogized applications of conventional intellectual property frameworks to safeguard data as an economic asset, regulators worldwide have started crystallizing bespoke sui generis data governance models
balancing access and control. This section analyzes key emergent approaches evidencing adapted data rights concepts [24].

1. European data governance act

The European Commission's proposed Data Governance Act (DGA) creates novel data sharing obligations upon companies meeting threshold criteria for "data altruism" services in the public interest. It defines categories of data permitting use for objectives like healthcare, combating climate change, improving mobility or facilitating official statistics. Registered non-profit data altruism organizations can request consent from businesses to share such data for stipulated public interest purposes. Refusals must be justified based on overriding legitimate interests. This innovative framework conceptualizes equitable public access rights over commercially held data deemed crucial for social welfare. It pivots data governance from proprietary control to stewardship norms oriented towards common good objectives counterparty rights [25].

However, critics argue reasonable interest exemptions to altruism remain loosely defined. Discretion to monetize data sharing permits encroaching commercialization. Requirements like anonymization and purpose limitation may still deter voluntary sharing of commercially sensitive data. Conflicts between DGA obligations and existing IP protections also remain unresolved. The DGA represents a landmark step towards a new social contract for data building on emerging notions of data trusts and collective pooled data governance. But translating principles like equitable access, stewardship duties and common good rights into operational frameworks warrants further policy evolution [26].

2. Data trusts

Data trusts comprise institutional structures pooling data rights and establishing collective controls over data access and usage terms on behalf of trust beneficiaries. They distribute oversight and bargaining power across stakeholders instead of consolidating proprietorship. Subject to data contributor permissions, trusts can facilitate controlled access by researchers, public agencies, businesses and civil society to fuel socially beneficial analysis while protecting against misuse. Different organizational implementations allow flexibility aligned to sectoral contexts. Public interest trusts steward data deemed crucial infrastructure for research and policy like environmental data. Consumer data trusts counterbalance the disproportionate power of tech platforms by granting collectives of users enhanced say over data sharing and monetization [27].

Community data trusts empower marginalized groups like indigenous people to secure equitable value from commercial data derived from their traditional knowledge. Employee data trusts proposal envisage workers pooling their personal data contributed to firms to gain collective leverage over access terms. Medical data trusts can enable sharing clinical data for research under ethics oversight. By
tempering unilateral commercial control, data trusts promise enhanced privacy protections, improved access for public welfare uses, and reduced algorithmic biases from unrepresentative data, stronger public oversight over technology risks, and equitable sharing of benefits [28].

However, critics argue that diffuse shared decision-making can hamper efficiency. Managing diverse user permissions on dynamic data presents logistical challenges. Power imbalances influencing control choices remain. And tensions between maximizing access versus minimizing harms persist, necessitating careful governance adaptation across domains. Additional policy support through data stewardship regulation can help incentivize adoption of bottom-up data trust mechanisms. Data rights in trust structures remain anchored in sui generis amalgamated consent by contributors rather than unilateral proprietorship. Implementing effective privacy and ethics safeguards tailored for distinctive datasets represents the key policy imperative. But the data trust paradigm offers promising possibilities for balanced data governance benefiting public and private interests [29].

3. Attribution rights over metadata

The exponential increase in data generation and circulation has made perceived attribution loss a growing concern for data producers. An emergent response gaining traction is asserting special data rights over metadata documenting provenance, traceability and usage history rather than restricting underlying data itself. This enables accrediting data originators while maximizing data access, thereby fostering attribution norms critical for scholarly communication, media integrity and public trust. The proposed intellectual property protection for data would grant sui generis rights to prevent unauthorized metadata removal rather than restricting data use or copying. Scholarly publishing entities have asserted claims over CrossRef DOIs and other metadata citing norms against misattribution [30].

News organizations advocate integrated rights protection over digital news metadata to combat plagiarism and misinformation. Commercial data vendors rely on end user licensing restrictions and cybersecurity measures for metadata-focused usage controls rather than blanket data IP protections. Technical mechanisms like digital watermarking which embed metadata directly within data presentations enable usage monitoring and enforcement against unauthorized removals. Blockchain-based data registries immutably log attribution transactions and commercial data licensing terms. However, policy tensions remain between proprietary metadata controls and public interest access needs. Overly restrictive terms may still impede data analysis even without constraints on underlying contents [31].

Ambiguities persist around public domain status of metadata like database
catalogs not subject to copyright. Additional sectors like healthcare require expanded access to metadata on data quality, integrity and provenance characteristics crucial for ethical usage decisions. Critiques about inherent uncertainties over accuracy and representativeness of metadata as a proxy for dataset reliability also warrant addressing through enhanced provenance standards [32]. The emergent attribution focused protection paradigms present a promising policy mechanism for incentivizing data production while safeguarding access. But further regulatory work remains in devising balanced frameworks and technology standards tailored across different sectors and avoiding proprietary overreach. The trajectory of augmented rather than absolute protection centered on accreditation and transparency norms merits continued analysis [33].

4. Technology-implemented usage controls

Beyond purely legal rights regimes, technological access control and enforcement tools provide alternate mechanisms for implementing data governance policies. These range from digital rights management (DRM) measures like encryption and access controls to immutable ledgers tracking data transactions. While technological controls face inherent limitations, in some contexts they offer pragmatic solutions for asserting bounded usage claims over factual data lacking inherent exclusivity. DRM tools commonly used for digital media enable similar technical restrictions on data usage, sharing and access. Encryption and permissions protocols grant selective access for stipulated purposes while preventing copying or extraction [34].

Digital watermarking imperceptibly encodes identifying metadata within data to monitor usage and enable claims over derivatives. Cloud access controls permit granular, context-specific conditional access privileges over data resources. While DRM faces hurdles like interoperability limits, spoofing attacks and public policy constraints on overriding access, thoughtful implementation in data contexts can enable differentiated access aligned with data sensitivity. Blockchain and distributed ledger technologies allow reliable tracking of data provenance and strong integrity protections. "Self-sovereign" identity schemes building on decentralized identifiers allow granular user-managed permissions. Blockchain-based registers can encode licenses, contracts and usage transactions pertaining to datasets [35].

However, challenges remain in reconciling transparency norms with privacy needs. Scalability and energy costs also require addressing. Ongoing experimentation, standardization and governance models tailored for varied data ecosystem needs hold promise in harnessing distributed ledger technologies as infrastructure for
trusted data sharing. Smarter integration of access control, monitoring and attribution technologies with formal IP-based data rights shows promise in implementing nuanced data governance capabilities. However standardization, interoperability and public oversight over proprietary DRM systems warrant attention. Holistic integration alongside contractual and regulatory policy mechanisms remains vital for balanced, ethical data governance [36].

IV. Discussion

While omnibus data regulations remain contested, context-specific data rights frameworks are emerging rapidly across sectors like healthcare, transportation, agriculture, smart cities, IoT and industrial data. Tailored sectoral rights address distinctive data sensitivities and innovation incentives. But variability risks fragmentation. The EU’s proposed Data Act defines rights over IoT data, granting users controls over data generated by smart devices. France's IoT data access provisions enable regulators to mandate data sharing. Germany's planned IoT register requires device specifics and contact information to facilitate safety oversight. Such sector-specific frameworks demonstrate sensitivities over IoT ecosystems vulnerabilities and lock-in effects from proprietary silos [37].

They highlight needs to balance device, network and cloud provider interests with consumer protection and public safety. In healthcare, clinical trial data access mandates like the US' FDAAA legislation and the EU’s policy 0070 requirements compelling companies to publish results within a year of marketing approval aim to protect public safety and research needs [38]. The UK's NHSX innovation arm has asserted special claims over NHS data to ensure public benefits. Such frameworks curb privatization of biomedical data. Automotive safety regulations increasingly require event data recorder information to be accessible for crash investigations. Some data sharing mandates like the EU's Cooperative Intelligent Transport Systems aggregate mobility data across vendors to enable safety and environmental applications [39].

Strategic data regulation initiatives exemplified by Germany's Data Strategy Law assert enhanced state privileges over data deemed crucial for national interests and competitiveness. Efforts are ongoing to secure community rights over data derived from traditional knowledge like genetic resources to ensure equitable benefit sharing. Models include protection frameworks like India's Biodiversity Act and the proposed MATRIX principles. Such context-specific data access and control provisions reflect distinctive public interests, ethics imperatives and innovation incentives tailored to sectoral data ecosystems. Although variability poses consistency challenges, bottom-up evolution enabling differentiated data governance merits analysis for insights on balancing complex multi-stakeholder equities [40].
Conclusion

Emerging data governance frameworks display continued conceptual tensions between public good knowledge sharing imperatives and private interests in extracting value from proprietary data control. This analysis of intellectual property theories as applied to conceptualizing rights over data reveals challenges in simply extending IP constructs centered on discretized expressions of creativity. Data possesses sui generis qualities of non-rival accessibility, continuous evolution, cumulative generation, and embedded factual character that warrant tailored policy accommodations. Neither unlimited proprietary monopolies nor total open access represent appropriate absolutes. Context-specific solutions balancing stakeholder equities prove necessary.

To stimulate socially beneficial innovation through fair data access while sustaining production incentives and ethical norms, regulatory approaches require careful calibration to sectorial contexts. Blends of public interest mandates limited but defined proprietary protections, collective oversight models and public-private governance rather than unilateral private rights over data resources appear promising. This necessitates further applied research evaluating data regulations using criteria like consistency, proportionality, flexibility, transparency, and accountability. Additional open questions include reconciling policy fragmentation across jurisdictions, addressing international data flows, bounding metadata controls, implementing effective remedies balancing deterrence and access, incubating data stewardship institutions, and specifying differentiation principles for variegated data.

Exploring collective licensing models and adapting insights across intellectual property domains also hold value. With data permeating innovation across industrial, governmental and scientific domains, intellectual property theories conceptualized for the sui generis data context can play a vital role in shaping national policies and international harmonization efforts. Balanced to equitably govern digital intelligence as a global public good while sustaining incentives for social welfare-oriented development. But significant analytical and policy innovation remains necessary for this emerging frontier at the intersection of law, ethics and technology.

References

4. Yulduz, A. (2023). Dealing with the Challenge of Climate Change within the Legal


