University of Pittsburgh School of Law

Scholarship@PITT LAW

2020

Tools for Data Governance

Michael J. Madison
University of PIttsburgh School of Law, madison@pitt.edu

Recommended Citation

Technology and Regulation

Tools for Data Governance

Michael J. Madison*

madison@pitt.edu

This article describes the challenges of data governance in terms of the broader framework of knowledge commons governance, an institutional approach to governing shared knowledge, information, and data resources. Knowledge commons governance highlights the potential for effective community- and collective-based governance of knowledge resources. The article focuses on key concepts within the knowledge commons framework rather than on specific law and public policy questions, directing the attention of researchers and policymakers to critical inquiry regarding relevant social groups and relevant data "things." Both concepts are key tools for effective data governance.

1. Introduction

Law offers no single or simple answer to the problems and opportunities afforded by data. For data scientists, commercial entities, and policymakers which may ask, "how should data be generated, or stored, or transferred, or used?," this article offers a short set of basic tools to use in developing suitable possibilities for governance and ethical practice. This is neither a detailed list of prescriptions nor an inventory or checklist of remedies for current controversies. Instead, the article offers two essential tools for imagining how to advance effective data governance. One consists of identifying and describing relevant social groups in which governance frameworks may be embedded. Two consists of identifying and describing relevant resources, or things, whose form and flow will contribute substantially to the welfare effects of the relevant data governance systems.

In part the aim of the article is to broaden relevant perspectives. Preparing the article followed a prompt to consider governing and regulating "data markets" relative to innovation, growth, and societal progress. That premise risks cutting off the inquiry prematurely. Markets, including regulated markets, are often too simplistic as descriptions of relevant problems or solutions, given what is almost self-evidently a complex challenge. State or government control or supply, as the usual alternatives to market regulation of problematic social phenomena, are likewise often too simplistic. Understanding data requires a broader view, adding the concept of commons governance to these two, in which "commons" embraces data sharing in some collectively managed or governed context. Data are almost always significant or valuable because they are shared.

In part the aim of the article is to provide a basic toolkit that is not tethered to immediate needs and that is adaptable and evolutionary in appropriate ways, as data governance questions challenge us to extend our imaginations. Some of this challenge is old. Along with researchers and industry, regulators and ethicists long ago began to confront the speed, breadth, and scale of the raw computing power now available at comparatively modest expense, so-called Big Data, and the rise of disciplines combined under the title "data science." Law and regulation have grappled with widely-deployed artificial intelligence (AI) systems, which feed on massive supplies of data.

What is new, and what calls for newly-flexible modes of thinking and practicing, is the apparent demise of human comprehensibility at the center of technology design and deployment. Computing speed, scale, and autonomous execution of networked computer systems today operate in ways that effectively embody *the absence of meaningful limits on the humans' capacity to discern patterns in data and to draw inferences from them.*

That concern is linked to virtually every area of human endeavor and more. Data undergirds both the "Internet of Things," material objects and environmental contexts in which networked sensors and actuators are embedded, and the "Internet of Bodies," in which connected devices are attached to or ingested by human beings.[1] The influences of data are seen in a growing number of techno-social systems, from manufacturing to health to politics.[2] One can imagine our data-saturated environment as a three-sided blend of the conceptual contri-

* Michael J. Madison is Professor of Law at the University of Pittsburgh School of Law. This paper was presented at the workshop 'Governing Data as a Resource' organized at Tilburg University in November 2019.

1 Andrea M Matwyshyn, 'The Internet of Bodies' (2019) 61 *William and Mary Law Review* 77.
2 Julie E Cohen, *Between Truth and Power: The Legal Constructions of Informational Capitalism* (Oxford University Press 2019); Brett M Frischmann and Evan Selinger, *Re-Engineering Humanity* (Cambridge University Press 2018).

butions of Claude Shannon as to information theory,[3] Alan Turing as to computability,[4] and Manuel Castells as to flows of power in the network society.[5]

To render these broader issues in more tractable terms, data governance asks more mundane questions: How do we get more data? Better data? More useful data? How do we control or limit data generation, or data distribution? How do we prevent or limit harms associated with data acquisition or retention? How do we increase, improve, or optimize social or economic value associated with data? How do we ensure that data are preserved appropriately, or made available for access appropriately?

Lurking close by are related questions about data governance in the context of specific sectors, industries, and fields. What is the role of data governance relative to personal privacy, employment, finance, national security, public administration, public safety, health and medicine, education, transportation, arts and entertainment, and more?

All the while, in almost all settings, understanding that we are sharing data, almost all of the time.

In sum, data governance must be able to accommodate both the broadest data-related questions asked above and also their context-specific applications. Because of that breadth, this article teases out arguments related to foundational questions of data sharing, rather than responding to the litany of questions just identified as "mundane" or sector-specific.

The article begins with a distinction between data form and data flow. This point is primarily descriptive. It has to do with what we focus on rather than simply on what we find. Both as a technical construct and as a social one, data appear to have a quantum character, in loosely metaphorical terms, meaning that data exhibit multiple and seemingly contradictory attributes. In any governance context, a critical and basic problem is: which attributes matter?

At times, data seem thing-like, a fixed object or objects capable of exclusive ownership and control and subject to regulation as if it were an artifact. That characterization of data-as-form seems most apt when data and datasets are subject to commodification and commercialization efforts.

At other times, and sometimes even at the same time, data seem wave-like, fluid, continuously evolving, even moving, aggregations of information that have power or effect by virtue of their scale or density on an ongoing basis rather than at a single moment. That characterization of data-as-flow seems most apt when data and datasets are parts of research programs and are put to other public uses.

In one sense data appear to be "private goods" and in another sense data appear to be "public goods,"[6] but that distinction can be overstated. Data are not always or necessarily "goods" of any sort.

The initial point is that the aims of data governance and regulation begin with exploring and describing both what data "is" and what data "ought to be," not in ontological terms, but in social terms, framed by data-as-form or data-as-flow. Section 2 expands on this.

That apparently simple distinction is fraught with complexity. Breaking down that complexity is the function of the rest of this article, in Sections 3, 4, and 5 below, describing a governance toolkit.

To render the toolkit comprehensible beyond the corridors and conference rooms of regulators and lawyers, the tools are conceptual rather than doctrinal. A conceptual approach avoids entanglement in disciplinary debates. In both descriptive and prescriptive senses, law has wrestled with the character of its basic approach to questions posed by knowledge and information, including data. One might start with issues of trade and commerce; or intellectual property and monopoly. One might focus instead on equity, autonomy, and dignity. A more integrative view would begin at a higher level of generality, asking whether regulatory challenges pose questions that are fundamentally private, including questions of contract (obligation) and tort, or fundamentally public, including questions of constitutional order and administrative law.

The article steers clear of such classification questions. It does not explore the details of specific legal systems or questions of legal rights and stakeholder interests. Instead, it situates questions of legal rule and governance strategy in the context of two distinctive concerns: what about groups, and what about things? Data-as-form and data-as-flow state two responses to a basic problem that data governance should address. It should address them, as an initial matter, by examining data governance as a species of institutional governance, and specifically knowledge commons governance. Section 3 below addresses that topic in greater detail.

The article likewise avoids undue reliance on the usual "either/or" questions that arise when law meets technology and when law meets information, such as individual rights vs. institutions and organizations, and/or the state. Security and stability vs. innovation and opportunity. Exclusivity vs. openness. And so forth. Those are proper governance concerns, and critically exploring groups and things helps us see how to advance them in specific and systematic ways.

But groups and things do something more. They open pathways into emerging research, scholarship, and (critically) experience that teach about a middle ground, in between markets and states, which is broad, useful, and too often overlooked, though it cannot be a panacea or a perfect solution. That middle ground is *knowledge commons*, which means social groups operating in structured ways relative to shared data.

Care must be taken with the language of commons and with what the language signifies. This is an argument for nuanced *governance* of data as a *shared resource* rather than for any hasty or wholesale abandonment of private interests, markets, or even the state. This is also an argument for an *ecological* and *evolutionary* perspective on data and data governance, a perspective that includes accounts of the roles of different actors, agents, and resources in producing both productive and unproductive outcomes of data-related systems. The word "commons" evokes precisely such a system-level perspective.[7]

The discussion of knowledge commons leads, in Sections 5 and 6, into the article's focus on two critical concepts: social groups, and things. These are high-level but nonetheless fundamental topics when investigating effective institutional governance of shared resources, such as data. And with those concepts, the article offers an introductory guide to fundamental data governance questions for the benefit of policymakers; institution and organization designers, builders,

3   Claude Elwood Shannon and Warren Weaver, *The Mathematical Theory of Communication* (University of Illinois Press 1998).

4   Charles Petzold, *The Annotated Turing: A Guided Tour through Alan Turing's Historic Paper on Computability and the Turing Machine* (Wiley Pub 2008).

5   Manuel Castells, *The Rise of the Network Society* (2nd ed. Wiley-Blackwell 2010).

6   Sabina Leonelli, 'Data — from Objects to Assets' (2019) 574 *Nature* 317.

7   Donella H Meadows, *Thinking in Systems: A Primer* (Diana Wright ed, Chelsea Green Publishing 2008).

and managers; and researchers and others who wish to find an initial hand-hold in this complex area.

2.   The Foundations of Data Pluralism

2.1   Data as Form, Data as Flow

It is a fiction that data "just is" (or "just are"), despite the fact that the word "data" itself derives from Latin for "given." Data are mined, produced, constructed, collected, prepared, cleaned, scrubbed, processed, analyzed, combined, sold, stored, and shared, all with explicit or implicit reliance on interpretive theories and models.[8]

Many metaphors appear in that sentence, some more helpful, some of them less so. All of them, in one way or another, suggest the static character of data. In that sense, data are things; or objects; or commodities. Data are fixed items and collections of information, documenting observations about the world. By implication data are scarce (metaphorically speaking) and valuable. Data-as-form captures the metaphorical instinct to treat data as things, or as a thing.

Metaphors are as inescapable in law as they are elsewhere in social life. By allowing us to describe one (less familiar) phenomenon in terms of another (more familiar) phenomenon, metaphors both describe our thinking processes and promote understanding. If we want to solve a problem, we must capture the problem in its full scope and character. At their best, metaphors are tools for doing that.

Yet metaphors are heuristics, and like all heuristics, they have their limitations and capacities to mislead. Data-as-form is, in this sense, incomplete.

One of the most popular umbrella metaphors for data is "the new oil." *The Economist*, a magazine, invoked that metaphor with the headline, "The world's most valuable resource is no longer oil, but data," alluding to the ubiquity of data, the quantity of data, its value as both commodity and as social and technical lubricant, and the associated economic value and market power of firms that deal in data.[9] The scholarly literature tends to join in the allusion.[10]

"Data as [the new] oil" can be misleading. Oil is tangible, and oil reserves are depletable. In most senses, data are intangible, and pools or collections of data are not depletable. More recently, *The Economist* has invoked a competing metaphor, "data as sunlight," signifying the fundamentally open character that data have, or should have.[11]

But some of the implications of the "oil" metaphor may be helpful. Oil is important and valuable partly because of its commodity character (oil in barrels rather than in untapped pools), but also partly because of its "infrastructural" qualities, in that it can be directed to numerous applications, with diverse value and values. Oil moves and flows, literally. Data are "flow" in related senses. Like oil, it is produced via complex technical processes. It can be "pooled" or

dis-aggregated. It can be a commodity itself. It lubricates social and technical process. It can be a vital component of numerous other technical and commercial applications. Data-as-flow captures the metaphorical instinct to look at data's fluid attributes.

Related tensions between data as form and data as flow are suggested by recent efforts by industry to clarify the meanings of metaphors such as "data lake" and "data warehouse" in describing modes of aggregating and managing data resources.[12] A "data lake" may combine data from multiple sources, suggesting flows of data; a "data warehouse" may organize data from a single source, suggesting a well-structured form.

These are not rigid characterizations. One should not be misled by the description of data in metaphorical terms. The key point, illustrated by the necessity of metaphor, is that data are *simultaneously* form and flow. No one, single, correct description of data exists on which we may ground some correct regulatory system. The present, massive moment in computing history, exposing the gap between human cognitive capabilities and computing capabilities, calls for intellectual and pragmatic humility and pluralism.

In that respect, it is important to amend the suggestion in the Introduction that data governance should build on systems perspectives on the origins and functions of data. Systems theory typically teaches a distinction between a resource system, sometimes referred to as a stock, and resource units, sometimes referred to as flows. The political scientist Elinor Ostrom, introducing her research on commons for natural resources, distinguished between fisheries and fish.[13] That distinction is most sustainable where biophysical attributes determine the identities and boundaries of the stock and the units. For data, biophysical attributes typically must give way to characterization and interpretation by humans, including differente modes of technology implementation . A systems perspective is still appropriate, even critical, as this article argues below. But identifying the relevant attributes of the system must be part of governance processes, rather than a lead-in to a governance processes.[14]

Three concrete contexts offer illustrations, before the article moves ahead to discussions of governance and resources more broadly, how current law, public policy, and practice rely on data-as-form and data-as-flow as fundamental framing devices. The illustrations are chosen because of the different respects in which they expose fundamental attributes of data in context. Here as elsewhere in this article, attention is drawn to concepts rather than to debates of the moment.

2.2   Copyright and Data

The first is copyright law. Both in the US and in Europe, data and databases as such are subject either to no copyright protection (data lie in the public domain) or to minimal or thin copyright protection. In the US, the Supreme Court opinion that holds that copyrightable works must reflect at least a modicum of "creativity."[15] Logically-structured collections of facts and data almost always do not. European copyright recognizes copyright in works that reflect the author's own

8   Sabina Leonelli, 'Data Governance Is Key to Interpretation: Reconceptualizing Data in Data Science' (2019) *Harvard Data Science Review* https://hdsr.mitpress.mit.edu/pub/4ovhpe3v accessed 7 February 2020.

9   *The Economist*, 'The World's Most Valuable Resource Is No Longer Oil, but Data,' 6 May 2017 https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data.

10   Dawn E Holmes, *Big Data: A Very Short Introduction* (Oxford University Press 2017); Viktor Mayer-Schönberger and Kenneth Cukier, *Big Data: A Revolution That Will Transform How We Live, Work, and Think* (Houghton Mifflin Harcourt 2013).

11   *The Economist*, 'Digital Plurality: Are Data More Like Oil or Sunlight?,' 20 February 2020 https://www.economist.com/special-report/2020/02/20/are-data-more-like-oil-or-sunlight.

12   Daniel E O'Leary, 'Embedding AI and Crowdsourcing in the Big Data Lake,' (2014) 29 *IEEE Intelligent Systems* 70.

13   Elinor Ostrom, *Governing the Commons: The Evolution of Institutions for Collective Action* (Cambridge University Press 1990), p. 30.

14   Christiaan Hogendorn and Brett Frischmann, 'Infrastructure and General Purpose Technologies: A Technology Flow Framework' (2020) *European Journal of Law and Economics* http://link.springer.com/10.1007/s10657-020-09642-w accessed 29 April 2020.

15   Feist Publications, Inc. v. Rural Telephone Service, Inc., 499 U.S. 340 (1991).

intellectual creation.[16] In practice that question is usually the "originality" of the work rather than the author's skill and labor. In cases involving collections of facts and data, originality has often been lacking.[17] The point is that modern copyright tends to advance a doctrinal judgment that data are best conceived in terms of data-as-form (is the work, as "thing," sufficiently original?) and that such a thing-like character is often absent. Although data often are human-created, data are and should be difficult to capture, control, and own, because of their obvious social value. Data might be form, but are not. In practice, as a consequence, data are flow.

Both the data-as-form and data-as-flow constructs can be modified by rule and by practice. Data producers and data controllers often have recourse to alternative legal strategies, both in commercial contexts and in research and government setting. Data-as-form approaches are observed in access controls imposed via contract and/or via technology limitations, as well as via legislative efforts to secure forms of exclusivity in databases that do not sound in copyright. The European Parliament, recognizing the poor fit between copyright and databases that is illustrated in the US by the *Feist* standard, adopted the so-called Database Directive in 1996. The Directive created a *sui generis* right to protect databases from appropriation, so long as the database in question represents a "substantial investment" of resources.[18] The inadequacies of that Directive have, in part, prompted the European Commission recently to propose a new "producer's right' in machine-generated data.[19] Data-as-flow approaches are evident in contract, technology, and commercial considerations combined in "Data as a Service," or "DAAS" arrangements. The categories are not rigid. The key is to see how they provide a conceptual foundation for the simultaneity of the conditions of day-to-day practice.

2.3   Public Health and Data

The second is law and public policy concerning public health and medical research. Data about individual health conditions and treatments is collected, abstracted, and generalized both in order to build predictive models of disease and contagion used for population-level interventions and to build diagnostic heuristics and predictive models used for individual-level interventions. In both settings, where models are built and interventions applied, data-as-flow defines the practice.

Where data are obtained or generated at the level of the individual patient or research subject, data-as-form may dominate. Data-as-form permit researchers and clinicians to describe the individual. Data-as-form permit them to document a collection of attributes about the individual. Data-as-form support policymakers and advocates, in contexts that highlight privacy considerations and human rights, who assert that, intuitively, the data "belong" to the individual because in some respects the data originated with or in that person. Commercial interests (and some research interests) claiming "ownership" of health-related data likewise invoke data-as-form arguments.

Legally, states have developed regulatory regimes to try to manage these conflicts, to protect the interests of researchers, the public, and commercial interests in generating better and more effective public

health and clinical medical strategies, and also to protect the interests of individuals in avoidable harm to interests in autonomy, privacy, and bodily integrity.

The US has done this via the Common Rule, a formal regulatory standard that governs ethical practice in biomedical and behavioral research involving human subjects, when that research is conducted (as almost all such research in the US is) with the support of federal funding or in federally-supported institutions. It provides that identifiable individual research subjects must give consent both to their participation in research and also to uses of associated individual data. In effect the Common Rule interposes strong initial data-as-form-based regulation on research programs animated by data-as-flow considerations.

Blends of data-as-form and data-as-flow may change. The Common Rule has now been changed. As of January 2019,[20] the Revised Common Rule substantially lowers the threshold for what amounts to "informed" content, meaning that research subjects no longer need to be provided with detailed and comprehensive information regarding uses to which "their" data may be put (quotation marks are included because, given the earlier discussion of copyright, the law may not support proprietary claims). It may be sufficient for researchers to disclose the simple fact that individual data may be shared. Data-as-form considerations are de-emphasized. Data-as-flow considerations are more prominent.

The illustration suggests both that neither data-as-form nor data-as-flow is necessarily superior in normative terms and also that the two framings may be combined, as in the copyright illustration earlier, in complex ways. Adoption of the Revised Common Rule was prompted by the power and potential of medical and public health research grounded in Big Data techniques, where sharing and combining data from multiple sources is increasingly the norm.[21] Critics point to alternative legal constructions, such as the European Union's General Data Protection Regulation (GDPR),[22] which blend individual patient interests and commercial interests differently.[23] The GDPR imposes significantly higher informed consent requirements with respect to storing and re-using individual health data. Normative assessment is complicated by additional data-as-form and data-as-flow attributes of US medical research systems. Authors of medical and public health research may be required by US law to share their research data by depositing data in public archives, a policy judgment based principally on data-as-flow.[24]

[16] Article 3(1) of Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases [1996] OJ L 77/20.

[17] Lionel Bently and Brad Sherman, *Intellectual Property Law* (4th ed. Oxford University Press 2014); C-604/10 Football Dataco v. Yahoo! UK and Others [2012] EU:C:2012:115.

[18] Article 7 of Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases [1996] OJ L 77/20.

[19] Peter K Yu, 'Data Producer's Right and the Protection of Machine-Generated Data' (2019) 93 *Tulane Law Review* 859.

[20] Dept. of Homeland Security et al., Federal Policy for the Protection of Human Subjects, 82 Fed. Reg. 7149, 7150/1 (Jan. 19, 2017).

[21] Willem G van Panhuis, Anne Cross and Donald S Burke, 'Project Tycho 2.0: A Repository to Improve the Integration and Reuse of Data for Global Population Health' (2018) 25 *Journal of the American Medical Informatics Association* 1608.

[22] 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 repealing Directive 95/46/EC (General Data Protection Regulation) OJ L [2016] 119/1.

[23] A Michael Froomkin, 'Big Data: Destroyer of Informed Consent' (2019) 21 *Yale Journal of Law & Technology Special Issue* 27; Lara Cartwright-Smith, Elizabeth Gray and Jane Hyatt Thorpe, 'Health Information Ownership: Legal Theories and Policy Implications' (2017) 19 *Vanderbilt Journal of Entertainment and Technology Law* 207.

[24] Deborah Mascalzoni and others, 'Are Requirements to Deposit Data in Research Repositories Compatible with the European Union's General Data Protection Regulation?' (2019) 170 *Annals of Internal Medicine* 332.

2.4   Biobanks and Data

A third illustration of diverse and changing uses of data-as-form and data-as-flow are biobanks, collections of biospecimens and related data derived from and used for research in both biomedicine and agriculture. Thousands of biobanks operate around the world storing tissue samples, genetic sequence information, and seeds, among other things. Their organizational structures are correspondingly diverse. Many are state-sponsored or supported. Some are private. Some are philanthropic. In cases where these enterprises collect and store data and physical specimens, as many do, their organizational design and governance and relevant legal regulation address data-as-form and data-as-flow perspectives in two layers.

One layer is biospecimens themselves, to which ethical, privacy, contractual, and tangible property interests may attach. They are data-as-form, in the sense of each biospecimen being a "thing," and a collection of biospecimens being a distinct "thing." Biospecimens are also data-as-flow, in that they are data as well as objects, and they have been collected and stored precisely because of their infrastructural, informational value to future researchers. A second layer is the informational data associated with the biospecimens, to which independent ethical, privacy, contractual, and *intangible* property interests may attach and which may have independent infrastructural importance for future research. The informational data are likewise data-as-form (the information associated with each specimen, and with a collection), and data-as-flow.[25]

3.   About Governance

Data-as-flow and data-as-form are rhetorical and propositional statements, but they are not pre-theoretical. They are not ontological statements about the true state of data. They are, by virtue of their metaphorical origins, judgments about the world, offered for their utility. They set out the initial conceptual vocabulary of this article. This Section provides the beginnings of its syntactical structure, which animates the analysis. If the challenge of data governance is identifying and advancing respects in which data-as-form should dominate data-as-flow, or the reverse, or neither, then how should that challenge be addressed?

This Section provides the first elements of a toolkit for analyzing situations and possibly recommending courses of action. It is a framework, which describes governance, institutions of governance, and the knowledge commons framework as an instrument for researching governance. Knowledge commons gets particular attention here because it provides as systematic framework for examining governance of shared knowledge resources, and because data governance is above all else, perhaps, a complex and sustained challenge in managing *shared* resources in institutional contexts.

Like a useful theory, a useful framework teaches us what conditions matter and what to look for, and why. As a device for assembling evidence, a framework should not be overly or prematurely precise and should initially accommodate multiple possible theories.[26]

3.1   Governance

The concept of governance is used here in the sense of collective or coordinated decisionmaking by individuals working together, about decisions on matters of collective interest. The emphasis on governance, rather than on law, regulation or public policy specifically or on coordination in the abstract, is based on and justified with respect to a fundamental anthropological instinct rather than a formal or positive legal one. Governance means individuals working together to form groups to solve their own problems.[27] A major thesis of this paper is that with respect to data, we should be asking about governance, not asking simply about law. Starting with governance opens the door to broader and more effective questioning about potential problems and solutions associated with data. Starting with markets or the state, per the Introduction, may pre-judge the character of both.

3.2   Institutions

Governance is best understood via its expression in institutions, rather than via the thoughts and behaviors of individuals. Individuals and their opportunities, thoughts, choices, and behaviors matter and, in a utilitarian sense, often matter most in final welfare judgments. But in practice, individual cognition and motivation are diverse. Efforts to understand governance primarily via references to an imaginary "model" human, responding to commands of the law, are destined to be unsatisfactory to the extent that the models do not match reality. This article foregrounds a framework that is grounded in empirics and pragmatics of institutions, meaning collections of individuals.

Governance is not limited, however, to formal institutions of the state, such as legislatures, courts, and administrative bodies. The reference to "institution" implies a broader view.

For a working definition of "institution," the article adopts the definition given by the economist Douglass North: the rules of the game of a society, devised by humans and shaping human behavior.[28] Also relevant, to similar if not identical effect, is the concept of the institution developed in modern sociology: institutions are stable behavioral patterns that reflect the coordinated behavior of individuals and organizations, where the relations define the actors rather than the other way around.[29]

The difference between the two perspectives, the former focusing more on rules that guide or determine patterned behavior, and the latter focusing on rules that reflect patterned behavior, is not determinative here. What matters is that institutions in either sense (or both senses) simultaneously produce and rely on well-understood sets of human-created norms to determine outcomes among a group of people who significantly self-identify with the enterprise in its own time. Groups may constitute and be denominated "communities" or "collectives" or firms or other enterprises. Membership or participation may be small or large. Group identity may be formally circumscribed or informal, dynamic, and fluid. Groups may exist in specific places and times, as firms or as cities, for example. They may combine mate-

25   Michael J Madison, 'Biobanks as Knowledge Institutions' in Timo Minssen, Janne Rothmar and Jens Schovsbo (eds), *Global Genes, Local Concerns: Legal, Ethical and Scientific Challenges in International Biobanking* (Edward Elgar Publishing 2019).

26   Elinor Ostrom and Michael Cox, 'Moving beyond Panaceas: A Multi-Tiered Diagnostic Approach for Social-Ecological Analysis' (2010) 37 *Environmental Conservation* 451.

27   Donald E Brown, *Human Universals* (McGraw-Hill 1991); Stuart P Green, 'The Universal Grammar of Criminal Law' (2000) 98 *Michigan Law Review* 2104.

28   Douglass C North, *Institutions, Institutional Change, and Economic Performance* (Cambridge University Press 1990).

29   Walter W Powell, 'Neither Market nor Hierarchy: Network Forms of Organization' (1990) 12 *Research in Organizational Behavior* 295; John Frederick Padgett and Walter W Powell, *The Emergence of Organizations and Markets* (Princeton University Press 2012).

rial and immaterial forms, transcending place and time in "imagined" communities of the sort described by Benedict Anderson.[30] Groups, loosely specified, are critical loci of governance in institutions.[31]

3.3   Institutional Governance of Resources

The rest of this Section offers a framework for investigating and understanding institutional governance of resources, including institutional governance relative to data, in ways that supplement the two usual sources of legitimate governance, states and markets.

Perhaps the most enduring and influential justification for the roles of markets and states in regulating resources, particularly relative to shared resources, is the story of the tragedy of the commons.[32] Modern researchers have come to identify the story closely with a well-known paper by the ecologist Garrett Hardin from 1968, but the story pre-dates Hardin's work.

The tragic commons offers a powerfully simplistic metaphor. As a result the story has been simultaneously a diagnostic tool, an explanation for historical developments, and a prescription. If resources are shared, they are likely to be over-exploited and ruined. To prevent the expected destruction, regulation should specify an actor or actors responsible for a defined set of resources, accountable either via the marketplace or via state mechanisms, and expect better results.

Legal scholars often have assimilated the tragic commons metaphor to problems in the creation and circulation of information and knowledge, such as production of inventions, new cultural works, management of data, personal information, and interests in privacy. The stereotypical implication is state supply of legal exclusivities in relevant intangibles, to be traded in private markets. Alternatively, the state may simply supply the resource itself, directly (by building and controlling it) or indirectly (by underwriting it). The expected solutions are intended to ensure that the resource exists in the first place, rather than over-exploited.

The tragic commons model works well in some settings. Positive law itself may at times be a resource that would not be adequately supplied absent state direction.[33] Various jurisdictions act differently on that institutional premise. US federal law is committed to the public domain. Other jurisdictions assert proprietary claims over the content of the law, in the name of the state. At best, in short, the tragic commons metaphor offers a helpful beginning. But its shortcomings are more significant. The inadequacies of the metaphor have been critiqued elsewhere at length. Only the briefest review is needed here.

In part, the tragic commons metaphor may mis-describe the resources themselves, particularly as to knowledge and information resources, such as data. The tragic commons metaphor typically posits a depletable resource. Even for tangible resources, that assumption may not hold. Material resources, even biophysical resources such as grazing pastures, may be regenerated or resupplied. For intangible and immaterial resources, such as data, consumption may

affect their value but not their existence. They suffer from no depletability problem. One significant problem is creating data resources in the first place, with the right attributes. Further resource-related questions are deferred to Section 5, below.

In part, and as relevant here, the tragic commons metaphor may mis-describe the actors involved. The tragic commons metaphor posits self-regarding, selfish decisionmaking actors with no means or motivation to acquire information about their neighbors' activities, no ability to plan for the future, no practice of coordinating their actions with their neighbors', and no capability for adaptation and innovation in the face of complexity.[34] The metaphor assumes no governance. Instead, it assumes a sort of pre-governmental, pre-political state of nature, with no background customs or rules regarding collective identity or appropriate behaviors, and primitive, one-dimensional individuals.

Obviously, the tragic commons metaphor is not intended generally to describe any actual world. But it may be taken as doing so, and when that happens, the metaphor may become something of a self-fulfilling prophecy. The failure of collective action that the metaphor predicts may provide a premise rather than a conclusion.

One may treat the production, consumption, and preservation of a shared resource as a challenge for collective action, rather than a failure of collective action. Can forms of collective action solve those challenges? Can those forms do so, particularly with respect to shared knowledge and information resources, in ways that are as welfare-enhancing as one supposes state production, distribution, and access?

3.4   Commons Governance

The path to a pluralistic modern understanding of institutional governance and the potential strengths of resource sharing institutions arose initially via the research of Elinor Ostrom. First collected in the 1990 book *Governing the Commons*,[35] the work of Ostrom and her colleagues, collaborators, and students carefully established, via an abundance of fieldwork and comparative analysis, that self-directed collaboration and collective action to solve resource management problems was possible – in practice, if not always in theory. Ostrom's adaptation of the "commons" framing not only enlarged policymakers' and scholars' fields of vision relative to shared resource challenge. This work re-introduced the idea of "commons" in an explicitly ecological sense, referring to actors, institutions, and resources interacting in systems in multiple interdependent ways.[36]

In *Governing the Commons* and later work, Ostrom added to economists' standard taxonomy of types of goods. Beginning with private goods (which area excludable and rivalrous), public goods (which are nonexcludable and nonrivalrous), and club goods (which are excludable but nonrivalrous, and sometimes referred to as toll goods), she added and focused on "common-pool resource systems," or "CPRs." CPRs are resources, rather than goods, a definition that expands their utility and functions to include uses beyond tradeability and consumption. CPRs are nonexcludable and shared but *depletable*, and subject to risks of overconsumption.

For common-pool resources, Ostrom described a series of considerations, or guidelines, indicating when informal systems of collective, community management of the resource was both feasible – contrary

30   Benedict Anderson, *Imagined Communities: Reflections on the Origin and Spread of Nationalism* (Verso 1983).
31   Michael J Madison, 'Social Software, Groups, and Governance' (2006) 2006 *Michigan State Law Review* 153.
32   Madelyn Sanfilippo, Brett Frischmann and Katherine Strandburg, 'Privacy as Commons: Case Evaluation through the Governing Knowledge Commons Framework' (2018) 8 *Journal of Information Policy* 116. A 'shared' resource is one that is produced, used, and/or consumed by multiple actors, either concurrently or sequentially.
33   Brigham Daniels, 'Legispedia' in Brett M Frischmann, Michael J Madison and Katherine J Strandburg (eds), *Governing Knowledge Commons* (Oxford University Press 2014).

34   Carol M Rose, 'Commons and Cognition' (2018) 19 *Theoretical Inquiries in Law* 587.
35   Ostrom, *Governing the Commons* (n 13).
36   Brett M Frischmann, 'Cultural Environmentalism and "The Wealth of Networks"' (2007) 74 *University of Chicago Law Review* 1083.

to the prediction of the tragic commons metaphor – and likely to generate sustained, welfare-promoting provision of that resource over time.[37] The word "commons" comes forward in this article from Ostrom's work. "Commons" means *not* fully open, unmanaged access to a resource, but instead collective institutional governance of a resource, embodying a set of strategies that solve coordination problems, known as social dilemmas. That mouthful of a phrase can be distilled into something shorter: commons means groups that engage in managed resource sharing. Institutional governance via groups may take the place of or exist in tandem with governance via exclusive rights and markets, (on the one hand) and governance via state provision or determination (on the other hand).

In highlighting the possible virtues of commons-based institutional governance of resources, Ostrom's work is important here in three respects.

One, Ostrom's guidelines for successful commons management have no direct or obvious utility in domains related to knowledge, information, and data. Virtually all of the research conducted for *Governing the Commons* and follow-on research focused on natural (i.e., biophysical) resources, such as water systems, forests, fisheries, and pasturage, which easily fit Ostrom's definition of a CPR. Though late in her career Ostrom and her colleague Charlotte Hess undertook some preliminary explorations of commons governance related to knowledge resources,[38] those efforts should be regarded more as encouraging further investigation rather than as definitive applications of Ostrom's work in new domains. Despite some preliminary efforts to apply Ostrom's work to data governance,[39] shareable knowledge, information, and data resources do not meet the definition of CPRs. In intangible, immaterial forms, knowledge resources are neither excludable nor depletable. Ostrom's commons guidelines should be set aside with respect to data governance. Whether and how collective- or community-based governance of data should function is a matter to be investigated afresh, via examining conditions in the field.[40] Ostrom's body of work exhibits a strong sympathy for collective self-determination and a strong skepticism of the role of the state, via formal property rights systems or otherwise. Those intuitions deserve empirical exploration in contexts related to data.

Two, Ostrom showed that understanding and developing effective institutional governance requires a strong dedication to empiricism and to comparative, contextual analysis.[41] Ostrom and her colleagues were motivated in part by specific resistance to the simplistic conceptual reasoning that is often associated with casual adoption of the tragic commons metaphor. In that spirit, Ostrom formalized her style of research in a strategy labeled the "Institutional Analysis and Development" framework (IAD) in order to support additional research.[42]

That style of analysis, if not that framework itself, is a critical step forward in understanding data governance.

Three, Ostrom highlighted the broad domain of successful resource governance strategies that rely neither on "market exclusivity" nor "state provision of a shared resource" (a strategy that would include a public policy declaring that a resource ought to be unowned and fully "open," as a part of a "public domain"). She titled the address she delivered in association with receiving the Nobel Prize *Beyond Market and States.*[43]

3.5   Knowledge Commons

The proposition that shared knowledge and information resources, such as data, ought to be subject to analysis and possible regulation via commons governance institutions of the sort just described, has been distilled into the knowledge commons research framework. That framework, described sometimes via the shorthand "GKC framework" after *Governing Knowledge Commons*, the title of the first volume of published knowledge commons research,[44] is an analytic tool motivated both by frustration with the tragic commons metaphor, as applied to information, and also by the strengths and style of Ostrom's research on commons. The GKC framework brings the ecological and systems spirit of that research into examinations of knowledge and information governance.

In contemporary research and policymaking, information production problems are simplistically modeled as overconsumption and free riding by multiple actors with access to a shared knowledge resource, leading to depletion and eventually to underproduction. Stereotypical solutions follow, modeled either as exclusive property rights transacted in markets (patents, copyrights), or as public goods provisioned by or underwritten by state authorities (such as scientific research). Problems of information privacy may be subject to equivalent stereotypical treatment, leading to proposals to vest strong exclusive privacy rights in individuals or to empower states to define privacy interests – to the exclusion of collectively self-directed privacy governance, in context.[45]

The GKC framework animates a research program intended to capture and inventory the domain of governance problems and solutions for knowledge and shared information resources. The GKC framework borrows its empiricism, its emphasis on context and setting, and its methodological pluralism from Ostrom's IAD framework. Similarly, the GKC framework anticipates the later development of one more theories or models of institutional design, individual motivation, and normative assessment. While the GKC framework is styled in the manner of Ostrom's IAD framework, it is not simply a special case of Ostrom's thinking or the IAD framework as such. Other scholars of information policy have similarly called for the development of governance strategies based on commons concepts: structured sharing.[46]

Clarifying the terminology helps to introduce the details of commons governance as a system by which some community or collective establishes and enforces principles of managed access to a shared resource. The underlying resource may be "purely" intangible and immaterial or a blend of material and immaterial attributes. The

37   Ostrom, *Governing the Commons* (n 13).

38   Charlotte Hess and Elinor Ostrom (eds), *Understanding Knowledge as a Commons: From Theory to Practice* (MIT Press 2007); Charlotte Hess and Elinor Ostrom, 'Ideas, Artifacts, and Facilities: Information as a Common-Pool Resource' (2003) 66 *Law & Contemporary Problems* 111.

39   Joshua B Fisher and Louise Fortmann, 'Governing the Data Commons: Policy, Practice, and the Advancement of Science' (2010) 47 *Information & Management* 237.

40   For the argument that Ostrom's instincts regarding governance, but not the details of Ostrom's program, should be applied to information privacy regulation, see Jane K Winn, 'The Governance Turn in Information Privacy Law' (2019) *SSRN Scholarly Paper ID 3418286* https://papers.ssrn.com/abstract=3418286 accessed 7 February 2020.

41   Brett M Frischmann, 'Two Enduring Lessons from Elinor Ostrom' (2013) 9 *Journal of Institutional Economics* 387.

42   Elinor Ostrom, *Understanding Institutional Diversity* (Princeton University Press 2005).

43   Elinor Ostrom, 'Beyond Markets and States: Polycentric Governance of Complex Economic Systems' (2010) 100 *American Economic Review* 641.

44   Brett M Frischmann, Michael J Madison and Katherine Jo Strandburg (eds), *Governing Knowledge Commons* (Oxford University Press 2014).

45   Sanfilippo, Frischmann and Strandburg (n 32).

46   Jorge L Contreras and JH Reichman, 'Sharing by Design: Data and Decentralized Commons' (2015) 350 *Science* 1312.

resource may be characterized by intellectual property rights and/or other exclusivity interests. The resource may originate in information that is characterized by no IP rights (public domain status). A *patent pool* and a *data pool* are both forms of knowledge commons, as the term commons is used here. A newsgathering and distribution collective, such as the Associated Press "wire" service, is a form of knowledge commons, although individual news stories are typically subject to few if any formal IP rights and in some countries, notably the US, are treated as presumptively open by virtue of constitutional requirements.[47] The relationship between the legal status of the underlying resource and the character of the resource management system is a question to be explored, not declared. Neither ownership nor openness is the end of the matter.

Commons governance includes a range of institutional governance practices under the "commons" umbrella. Because knowledge and information resources may be defined and regulated by positive law, commons governance systems and market-based systems and formal state regulation may be linked and overlap in specific contexts. Further, no bright line exists to divide *knowledge* commons, which are directed primarily to information resources, from other sorts of commons, such as natural resource and environmental commons studied by Ostrom and her colleagues, and urban commons, which refer to governance of urban planning and design.[48] The acronym CPR, which in social science research refers to "common-pool resource," also appears in property law theory as "common property regime," a commons-like governance system anchored in analyses of infrastructural resources such as roads. Common property regimes highlight increasing returns to scale as more and more people consume a resource of a given size.[49] Infrastructural resources, because of their shared character, are often governed as commons.[50] The practice of "commoning" usually refers to politically or ideologically-motivated practices combining local resource governance institutions and self-directed community governance.[51]

The details of the GKC framework as a research instrument are described elsewhere.[52] The key insight of the framework is not whether the institution "is" or "is not" a commons. Rather, the question answered by the framework is whether and how some knowledge or information resource is governed as a *shared resource* via some community or collective, as an alternative to knowledge governance in markets, founded on claims of exclusivity of right, such as patents or copyrights or to knowledge governance via state intervention, provision, or subsidy. Commons governance systems may play important roles with respect to market-based and government-supplied resources. The question is whether some knowledge or information resource presents, in substantial part, hallmarks of

structured sharing.[53]

Knowledge commons governance is neither rare nor novel, nor is it limited to specific economic or cultural niches, such as small communities. The GKC framework supplies a means of describing the breadth of the field in a systematic way. The functionality of durable knowledge commons governance – broadly across technical and cultural domains, at different scales, and in specific cases – has been demonstrated in cases across a diverse range of contemporary and historical settings, including both technology development and cultural creation.[54] Janis Geary and Tania Bubela provide an exemplary case study of knowledge commons in a specific and focused case of contemporary life sciences research.[55] Knowledge commons has been used to analyze the field of microbial biology.[56] The GKC framework is consistent with research on patent pools, open source software development, and clearinghouses[57] and other institutions for collective governance of shared resources, including data and datasets. These have been documented in historical settings,[58] in less developed countries,[59] in large-scale, critical scientific and health related research networks,[60] in large scale commercial settings,[61] and in Big Data-enabled scientific research communities.[62]

3.6   Rules and Norms

The GKC framework is primarily descriptive, rather than normative. It aims to surface attributes of institutions via examination of specific cases for potential comparative assessment, using tools borrowed in part from social science, in part from the humanities, and in part from law. (The framework is intended to be accessible to and usable by researchers from each of these domains.) Users of the framework and students of knowledge commons research often focus on the systems of formal and informal rules, norms, customs, and practices by which communities and collectives govern themselves and govern relevant resources. In GKC research as in much of Ostrom's work, these are "rules in use," signifying their empirical rather than normative status. For purposes of comparative institutional analysis, these rules in use may be productively compared with rules and norms in evidence in market-based governance systems and those prescribed

47   Michael J Madison, Brett M Frischmann and Katherine J Strandburg, 'Constructing Commons in the Cultural Environment' (2010) 95 *Cornell Law Review* 657.
48   Sheila R Foster and Christian Ianone, 'Ostrom in the City: Design Principles and Practices for the Urban Commons' in Blake Hudson, Jonathan Rosenbloom and Dan Cole (eds), *Routledge Handbook of the Study of the Commons* (Routledge 2019).
49   Carol M Rose, 'The Comedy of the Commons: Commerce, Custom, and Inherently Public Property' (1986) 53 *University of Chicago Law Review* 711.
50   Brett M Frischmann, *Infrastructure: The Social Value of Shared Resources* (Oxford University Press 2012).
51   David Bollier and Silke Helfrich (eds), *Patterns of Commoning* (Common Strategies Group 2015).
52   Michael J Madison, Brett M Frischmann and Katherine J Strandburg, 'Knowledge Commons' in Blake Hudson, Jonathan Rosenbloom and Dan Cole (eds), *Routledge Handbook of the Study of the Commons* (Routledge 2019).

53   Frischmann, Madison and Strandburg (n 44).
54   Frischmann, Madison and Strandburg (n 44); Katherine J Strandburg, Brett M Frischmann and Michael J Madison (eds), *Governing Medical Knowledge Commons* (Cambridge University Press 2017).
55   Janis Geary and Tania Bubela, 'Governance of a Global Genetic Resource Commons for Non-Commercial Research: A Case-Study of the DNA Barcode Commons' (2019) 13 *International Journal of the Commons* 205.
56   JH Reichman, PF Uhlir and Tom Dedeurwaerdere, *Governing Digitally Integrated Genetic Resources, Data, and Literature: Global Intellectual Property Strategies for a Redesigned Microbial Research Commons* (Cambridge University Press 2016).
57   Geertrui van Overwalle (ed), *Gene Patents and Collaborative Licensing Models: Patent Pools, Clearinghouses, Open Source Models, and Liability Regimes* (Cambridge University Press 2009).
58   Tine de Moor, 'From Historical Institution to Pars Pro Toto: The Commons and Their Revival in Historical Perspective' in Blake Hudson, Jonathan Rosenbloom and Dan Cole (eds), *Routledge Handbook of the Study of the Commons* (Routledge 2019).
59   Jeremy De Beer and others (eds), *Innovation & Intellectual Property: Collaborative Dynamics in Africa* (Published by UCT Press in association with the IP Unit, Faculty of Law, University of Cape Town and Deutsche Gesellschaft für Internationale Zusammenarbeit 2014).
60   Amy Kapcynski, 'Order without Intellectual Property Law: Open Science in Influenza' (2017) 106 *Cornell Law Review* 1593.
61   Henry Chesbrough, *Open Innovation: The New Imperative for Creating and Profiting from Technology* (Harvard Business School Press 2003).
62   Michael J Madison, 'Commons at the Intersection of Peer Production, Citizen Science, and Big Data: Galaxy Zoo' in Brett M Frischmann, Michael J Madison and Katherine J Strandburg (eds), *Governing knowledge commons* (Oxford University Press 2014).

in state-based regulatory settings. For students and practitioners of data governance, the intuitive answer to "how should we regulate?" takes the form of "these are the appropriate rules."

In the context of knowledge governance, the temptation to prioritize examination of the rules, empirically or normatively, may be resisted. It risks putting the proverbial cart before the horse. The review of commons governance shows why: commons governance is collective management of a shared resource by or in a group. The role of the collective is largely to define its own governance system relative to dilemmas associated with specified resources, producing a form of institutional governance in context. This article has described the fundamental problem of governing data sharing in terms of two conceptions, data-as-form and data-as-flow. It argues next that understanding data governance should begin not with the rules, but instead with two key phenomena: groups and things.

4.   About Groups

"Groups" means formal and informal collections of people, who identify themselves with the group (perhaps closely, perhaps loosely, and perhaps in variable numbers over time) and who adopt and enact practices that are aligned with the interests and identities of the group. When knowledge commons governance research refers to institutional governance of shared resources by self-directed communities and collectivities, it refers to groups solving social dilemmas regarding those resources. *Beyond Markets and States*, the title of Ostrom's Nobel Prize address, is read fairly to claim that governance by groups is an empirically valid mode of resource management.

In practice, that summary opens at least three key lines of inquiry as conceptual matters.

The first is the most pragmatic: In a resource governance context, does one or more groups exist that might serve as governance vehicles? How might such a group be identified, defined, and organized? Should law or regulation be invoked to motivate or to discourage group formation as part of an institutional governance strategy?

The second concerns the possible governance contributions by groups. Groups might generate relevant rules, norms, and practices on their own, such as a voluntary association, or might serve as agents for administering and enforcing rules and norms generated elsewhere, such as employees of a for-profit firm. Groups might serve as collective institutions in a cognitive sense, so that the collective is able to identify and act on information that is not equally accessible or useful to individuals acting alone. Groups might act as loci for interpretive practices by which society gives shape and meaning to places and resources, as suggested in multiple traditions of Science and Technology Studies. In each of these respects, where present, groups may participate in resource governance practices.[63]

The third and most important here concerns ways in which groups may be anchors for two especially critical conceptual foundations for institutional governance of shared resources: *trust* and *polycentricity*. Data governance strategies should explore both.

4.1   Trust

Trust represents the sense that trust mechanisms are critical to cooperative arrangements.[64] It also represents the sense that actual human beings have greater capabilities for understanding and adapting to complex social and environmental challenges, and for doing

so cooperatively rather than being coerced to do, than is predicted by tragic commons metaphors and presumptions of selfish behavior following the pursuit of rational self-interest.[65] Trust may operate bilaterally, between individuals or between an individual and an institution. Trust also operates critically among populations of individuals and an institution. For governance by groups, social trust mechanisms must operate at some level among the members of the group, relative to one another and relative to the purposes of the group. It has been argued that trust generally consists of means by which individuals cope with the fact that others may exercise their own freedom.[66] But no single, optimal definition of trust exists.

Likewise, no single social or policy mechanism works universally to promote trust and promote group formation, identity, durability, or adaptability, or to undermine trust or to prevent it from forming. Group-based resource governance may be unhelpful or harmful, or may create unmanageable conflict with other governance institutions. The research literature on trust and cooperation is vast, and it covers sociological, anthropological, economic, political science, and philosophical domains.[67] Emphasizing reciprocal relations between community members, for example, is sometimes suggested as a critical ingredient in effective cooperative settings, an idea that may be traced back to early work on gift economies. But the details matter. "Pay it forward" reciprocity strategies may be as important to trust formation as "pay it back" strategies, or more so.[68] Trust creation and reinforcement may depend on relationships among group decision-making rules (such as enforcement norms, or exit/entry criteria) and the development of shared collective identity (such as "who we are" questions).

This makes trust an ecological and structural question as well as a matter of individual cognition.[69] The research and policy challenge is to design and support institutions where the benefits of individuals' cooperative capabilities can be put to good use, where shared resources can be governed effectively, and where the weaknesses of a trust-based model are minimized. Cooperative capabilities are unevenly distributed, for example, and trust mechanisms may be riddled with harmful power and influence dynamics. Trust is itself, significantly, a shared resource, and governance of that resource is likely necessary as part of broader resource governance strategy.

4.2   Polycentricity

That trust is a shared resource subject to governance, as part of governance of a shared knowledge resource such as data, points to the idea that governing groups may overlap and intersect. *Polycentricity* captures that concept, in the sense that any institutional design for governance is likely to be most effective when it is characterized and implemented in a decentered way, with multiple loci of authority and responsibility, rather than a single center of regulatory agency, intersecting with one another at different scales[70] and relying on individuals' diverse motivations for participating.[71]

[63] Madison, 'Social Software, Groups, and Governance' (n 31).

[64] Kenneth Joseph Arrow, *The Limits of Organization* (Norton 1974).

[65] Bo Rothstein, *Social Traps and the Problem of Trust* (Cambridge University Press 2005).

[66] Niklas Luhmann, *Trust and Power* (English edition, Polity 1979).

[67] Diego Gambetta (ed), *Trust: Making and Breaking Cooperative Relations* (B Blackwell 1990).

[68] Toshio Yamagishi and Karen S Cook, 'Generalized Exchange and Social Dilemmas' (1993) 56 *Social Psychology Quarterly* 235.

[69] Kenneth W Abbott, Jessica F Green and Robert O Keohane, 'Organizational Ecology and Institutional Change in Global Governance' (2016) 70 *International Organization* 247.

[70] Julia Black, 'Constructing and Contesting Legitimacy and Accountability in Polycentric Regulatory Regimes' (2008) 2 *Regulation & Governance* 137.

[71] Yochai Benkler, 'Law, Innovation, and Collaboration in Networked Econo-

Those multiple centers may be informal or formal or blends of the two. Groups may be organized hierarchically. Smaller groups may be "nested" within a larger group. Groups may be linked to on another in a network of distinct and/or overlapping nodes of different scales. Polycentric systems can be flexible and adaptable across time, scale, and community form. They can support enforcement and accountability mechanisms at different scales, enhancing legitimacy, accountability, and administrability of governance systems as a whole.

So, just as trust is a key governance variable to be explored, polycentricity does not solve all problems. One must still carefully consider the scope of authority and its mechanisms of accountability and legitimacy. Like all governance systems, and like trust, polycentric systems are subject to appropriation and abuse via dynamics of power, wealth, and status. Polycentricity is not a cure-all. It is an analytic strategy, and polycentric systems can be made stronger and weaker.[72]

4.3   Groups and Data

Group-based perspectives, including polycentric governance and emphasis on structures that both generate and rely on social trust, are consistent with but perhaps more nuanced and potentially effective than other norm-based approaches that are not so explicitly pluralistic. Governance of shared data resources with reference to groups helps us organize possible strategies distinguished as data-as-form and data-as-flow. The absence of relevant groups relative to those data resources suggests a different range of strategies distinguished along those lines. For example, certain approaches to "open" data governance (a species of data-as-flow) may be better appreciated and have greater impact if described as parts of polycentric governance, including "best practices" recommendations; "fair practices" approaches, such as the Fair Information Practice Principles (FIPP) for personal data collection and the FAIR Data Principles for scientific data management; suggestions that all of data or all of knowledge constitutes a single, global shared resource;[73] and advocacy under labels such as Open Science and Open Data. In these contexts, "openness" and "fair" practices are achieved by paying careful attention to institutional attributes of groups and fields.[74]

Historical data governance practices are similarly illuminated by prioritizing questions about groups, trust, and polycentricity. The historian Will Slauter argues persuasively that seventeenth century English publishers strategized ways to obtain exclusivity in shipping and price information (data-as-form).[75] Modern copyright and its near-total exclusion of data from legal ownership is in many respects a product of those strategies, their modern analogs, and resistance by other groups in UK and American legal systems (data-as-flow). The political scientist James Scott suggests, provocatively, that central state authority exists not only to enhance the well-being of citizens but to

render information about them "legible," as data (data-as-form).[76] He argues that alternatives to the modern state, in localized, collective self-governance, may be equally effective at promoting well-being and offers the benefit of maintaining critical distances between the state and its subjects (data-as-flow).

5.   About Things

"Things" captures a broad range of related phenomenon: items, units, commodities, embodiments, objects, artifacts, and stuff, both material and immaterial, analog and digital. With such a broad beginning, semantics and ontologies can get tricky, and interpretive techniques must be developed to sort out relevant distinctions.[77] One object may embody more than one thing, and one thing may be embodied in more than one object. A "work of art" such as a novel may embody a distinct "work of authorship" or "copyright work"; that copyright work may be embodied in numerous copies of the novel. One thing, such as a the novel, may be part of another thing, such as a library, and may itself consist of other things, such as literary elements, and chapters. Identity is another concern. In the larger collection, the smaller unit may be separable. But not always. A gallon of water poured into a river mixes inseparably with the rest of the river. A gallon of water can be extracted from the river, but that gallon is not the same gallon as the water previously poured in. Origins, possession, and authenticity also shape the definitions, meanings, and purposes of things. Things are often associated with specific individuals. They are also often associated with social groups.

The word "thing" is a broad and inclusive way to refer to "resource," as that word and concept have contributed earlier to discussions of governance. When knowledge commons governance research refers to institutional governance of shared resources by self-directed communities and collectivities, it refers to groups solving social dilemmas regarding the creation, use, and preservation of things, treating things as a flexible category that allows researchers and analysts to explore widely.

In practice, that summary opens at least two key lines of inquiry as conceptual matters.

The first returns to the prompt with which the article began: the essential distinction between data-as-form and data-as-flow. That distinction suggests asking, foundationally, what is a thing, and how do we know? Whereas the last Section built conceptually on the contributions of Elinor Ostrom, to a sizable degree this Section moves beyond Ostrom. Ostrom's work on institutional governance and commons typically relied heavily on analysis of natural resources, which come to us with given and mostly unmodifiable biophysical attributes. Ostrom's later work, on knowledge, tended to treat "knowledge" as a single, undifferentiated resource. Neither approach suits the GKC framework. Neither approach suits data.

The second concerns relationships between groups and things. Those relationships are often fundamentally ecological and systemic. The social groups that construct and manage resources may be produced, reinforced, and reproduced by the identity of the resource and by the group's governance practices relative to the resource, both as to the internal dynamics of social groups and as to relationships between social groups.[78] How should those relationships be explained?

---

my and Society' (2017) 13 *Annual Review of Law and Social Science* 231.

72 Black (n 70). The next Section offers a parallel point regarding resource systems themselves, which can be designed flexibly to operate at greater or lesser scales. The cognitive scientist Herbert Simon characterized organisms with this character as "nearly decomposable." He argued that the "decomposability" strategy for managing adaptation in complex environments rendered such organisms particularly fit in evolutionary terms. HA Simon, 'Near Decomposability and the Speed of Evolution' (2002) 11 *Industrial and Corporate Change* 587.

73 Hess and Ostrom, 'Ideas, Artifacts, and Facilities: Information as a Common-Pool Resource' (n 38); Jennifer Shkabatur, 'The Global Commons of Data' (2019) 22 *Stanford Technology Law Review* 354.

74 Liz Lyon, Wei Jeng and Eleanor Mattern, 'Research Transparency: A Preliminary Study of Disciplinary Conceptualisation, Drivers, Tools and Support Services' (2017) 12 *International Journal of Digital Curation* 46.

75 Will Slauter, *Who Owns the News? A History of Copyright* (Stanford University Press 2019).

76 James C Scott, *Seeing like a State: How Certain Schemes to Improve the Human Condition Have Failed* (Yale University Press 2008).

77 Michael J Madison, 'IP Things as Boundary Objects: The Case of the Copyright Work' (2017) 6 *Laws* 13.

78 John Seely Brown and Paul Duguid, *The Social Life of Information* (Harvard

5.1   What is a Thing

Identifying and defining things are problems in epistemology that go back to Aristotle. The question here is not so broad. The question is governance: what are the things that form parts of governance systems? What are resource systems, and what are resource unit? Borrowing the concept of polycentricity, how do multiple resource systems interact, overlap, and align? Where do relevant resources come from; how are relevant resources used, consumed, and applied; and how, if at all, are relevant resources preserved over time? Data-as-form and data-as-flow are then both inputs into governance analysis and outcomes of governance analysis.

For biophysical resources, answers to most of these questions may be relatively straightforward; resources are the objects of governance. For knowledge and information resources, including data, resources are both subject and objects of governance. Governance often creates (produces, consumes, preserves) the things to which governance applies. Prioritizing things in governance is a way of prioritizing a key set of critical questions. Pragmatically, a critical perspective on governance means that little turns on classifications of things resources as inherently private goods, public goods, club or toll goods, or common-pool resources. The tools of law and policy as well as the experiences of social life teach that boundaries and classifications among these categories can be modified in many settings, disrupting what otherwise might be standard prescriptions based on the logic that gives priority attention to commons tragedies. A functional approach, based on an empirical approach to ecologies in practice, is preferred.

Data depend on their reference and relationships to underlying phenomena. In that sense, data are evidence of something else.[79] They are, almost by definition, both things in themselves and also versions of something else. Data signify a problem long recognized in mathematics, computer science, geography, and literature: to be useful, a model or map must stand in for the whole but not be identical to it.[80] Data are sometimes characterized as "raw" or "cooked," a metaphorical framing that suggests the degree to which data directly (raw, unprocessed) or indirectly (cooked, processed and analyzed) relate to their source. The metaphor departs from its partial origins in the anthropological literature, as a reference to the construction of conceptual oppositions.[81] But the allusion gets at something equally fundamental. Both the identity and the attributes of data, databases, and datasets, including attributes implicating exclusivity and shareability, are matters of design as well as physics or economics.

5.2   Things and Groups

Significantly, social groups are among the most fundamental "designers," even with respect to such traditional resources as property in land. The legal historian Molly Brady, for example, has carefully documented that the historical meaning of the phrase "metes and bounds" in the law of real property refers to boundaries identified by local social and community practices, rather than to fixed boundaries specified by surveyors.[82]

In domains of knowledge and information, including data, the absence of a standard or uniform material reference (unlike land) means that the role of social relationships in constituting things and resources, both in social life and in legal processes, is both broader and deeper.[83] The argument draws on research in Technology Studies and Information Science, rather than legal scholarship.[84] Scholars have researched access to immaterial goods;[85] have explored governance of resources that generate additional resources (so-called "generative" phenomena);[86] and explored modern technologies such as open source computer programs, in which the group and the object constitute each other.[87]

Observing that things may be constructed socially, particularly for purposes of governance, does not imply that those processes of construction are simple or straightforward. (Nor does it imply that material objects do not have a physical reality.) The variability and complexity of those processes; the possibilities that they may or may not be linear and/or purposeful; the fact that they likely involve multiple social systems, including law; and the reality that individual actors in those systems, even within social groups, may have conflicting motivations, are precisely what give rise to the need to examine those processes critically.[88]

In commercial law settings, for example, two actors may agree by contract to treat a dataset as a tradeable commodity even while formal IP law considers that same information to be unowned and unownable. Customary practices in many fields construct domains of things for disciplinary purposes, such as the "copy" that has been the unit of text for both publishers and journalists. For public policy reasons, legal institutions may declare an absence of thing-like character, in order to deprive others of the power to claim property-like exclusivity in them. Patent law resists granting exclusive rights in laws of nature and abstract ideas. Property scholars who are committed to the central role of "things" in property law have begun to explore the legal "toolkit" of doctrines and arguments needed to construct property resources at different scales, producing an architecture of property things.[89]

As noted earlier, the GKC framework for researching knowledge

Business School Press 2000); Madison, 'Commons at the Intersection of Peer Production, Citizen Science, and Big Data: Galaxy Zoo' (n 73); Thomas C Schelling, *The Strategy of Conflict* (Harvard Univ Press 1960); Susan Leigh Star and James R Griesemer, 'Institutional Ecology, `translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39' (1989) 19 *Social Studies of Science* 387.

79   Christine L Borgman, *Big Data, Little Data, No Data: Scholarship in the Networked World* (MIT Press 2015).

80   Brian Cantwell Smith, 'The Limits of Correctness' (1985) 14,15 *ACM SIGCAS Computers and Society* 18.

81   Claude Lévi-Strauss, *The Raw and the Cooked: Introduction to a Science of Mythology* (Penguin Books 1992).

82   Maureen E Brady, 'The Forgotten History of Metes and Bounds' (2019) 128 *Yale Law Journal* 872.

83   Michael J Madison, 'Law as Design: Objects, Concepts, and Digital Things' (2005) 56 *Case Western Reserve Law Review* 381.

84   Geoffrey C Bowker and Susan Leigh Star, *Sorting Things out: Classification and Its Consequences* (MIT Press 1999); Bruno Latour, *Reassembling the Social: An Introduction to Actor-Network-Theory* (Oxford University Press 2007); Henry Petroski, *The Pencil: A History of Design and Circumstance* (Knopf 2006).

85   Jessica C Lai and Antoinette Maget (eds), *Intellectual Property and Access to Im/Material Goods* (Edward Elgar Publishing 2016).

86   Jonathan Zittrain, *The Future of the Internet and How to Stop It* (Yale University Press 2008).

87   Christopher M Kelty, *Two Bits: The Cultural Significance of Free Software* (Duke University Press 2008); Charles M Schweik and Robert C English, *Internet Success: A Study of Open-Source Software Commons* (MIT Press 2012).

88   Ellen P Goodman (ed), *The Atomic Age of Data: Policies for the Internet of Things* (Annual Aspen Institute Conference on Communications Policy 2015).

89   Thomas W Merrill and Henry E Smith, 'The Architecture of Property' in Hanoch Dagan and Benjamin Zipursky (eds), *Research Handbook on Private Law Theories* (Edward Elgar Publishing Forthcoming) <https://papers.ssrn.com/abstract=3462643> accessed 7 February 2020; Lee Anne Fennell, *Slices and Lumps: Division and Aggregation in Law and Life* (University of Chicago Press 2019).

commons emphasizes how social groups develop governance to address social dilemmas, or problems in cooperation.[90] Analysis of social dilemmas in complex settings may be simplified somewhat by techniques of "decomposing" large systems into small components.[91] Larger things may contain small things.

In sum, as to the identity of relevant resources, the possible absence of linearity and the importance of context should be emphasized, and over-reliance on *ex ante* categorization should be avoided. That point has particular significance with respect to data. Modern research demonstrates how scientific research consists of reciprocating processes rather than a progression from "basic knowledge" to "applied knowledge," including technology development and commercial application.[92] Likewise, research data production and management is now likewise often expressed in cyclical terms.[93] Data are sometimes characterized entirely as an infrastructural resource.[94] That focus highlights the many ways in which data use creates spillovers in multiple fields, in both expected and unexpected ways. But that infrastructural designation should be taken only as the beginning of an examination of appropriate governance, because infrastructure is a designed and socially constructed resource much as any other knowledge or information resource is.[95]

5.3   Things and Data

One strength of the word "resource" is that it properly evokes relationships between resources in resource systems or ecologies. Awareness of data ecologies for governance analysis aligns specifically with the emphasis that the GKC framework places on governance in broad context. An ecological perspective requires examining interdependencies between those resources and related resources, as systems, involving both immaterial and material attributes and evolution and variations across scales.

Understanding ecologies of data "things" should take account of the data collection and management practices associated with Big Data, with special attention given to the sources of the now-standard "three v's" of Big Data (volume, variety, and velocity), all the way down to hand-curated data collections. Different settings, resources, and resource systems may call for different governance judgments as to relevant social groups and as data-as-form and data-as-flow considerations.

Those settings and resources may include the following. The classification below is crude. Many overlaps exist among tools, products, services, and research outputs, and multiple opportunities exist to deploy characterizations of data-as-form and data-as-flow.

- Techniques and technologies for observation, experimentation, data collection, association, and construction of databases and datasets. These may include physical devices (the Internet of Things and the Internet of Bodies) as well as digital protocols, including computer programs, data formats, and other digital standards) for sensing and observing, for data transmission and communication, and for creating and managing the resulting data collections. MapReduce is an example of a digital computing par-

adigm for managing super-large datasets in a distributed computing environment.[96]

- Processes and systems of data stewardship, which emphasize cleaning, scrubbing, normalizing, manipulating, classifying, and maintaining data for storage, analysis, use and application.[97] Data ontologies, data schema, and data storage techniques and models are critical to ensure both technical synthesis and interoperability where data from multiple sources are brought together for use as shared resources, as in data repositories or other data infrastructures.

- Analytics, interpretations, and applications. These occupy an enormous analytic space in their own right, because "data" as governance subjects overlap with "algorithms," "AI," and "platforms" as technologies and institutions for data mining strategies; pattern analysis; and services, products, and new knowledge forms built on those patterns, as governance subjects. As machine learning technologies enable the automatic adjustment of data collection practices via embedded sensors, boundaries blur between data and AI. So-called smart machines learn from old data and collect new data differently. Data visualization tools are critical here, as are conceptual maps and models.[98]

The worlds of data may be changing and expanding so quickly, and this three-part division of data-related resources may be so imprecise, that it may seem unwise to advance the concept of things as a key governance concept. Yet two brief examples illustrate how focusing on things in governance, and particularly in commons governance of shared data, can illuminate specific data-related challenges.

A first example comes from outside the law, in coordination challenges among social groups within a given broad field. Academic researchers know this as the problem of coordinating across research disciplines. Because so much scholarly research now centers on data along with disciplinary knowledge, researchers confront new governance challenges even within institutions long associated with openness and sharing, such as scientific communities and research universities. The knowledge sharing norms of medical researchers overlap with but are also distinct from knowledge sharing norms of engineering researchers and social work researchers, for example. Data-as-form and data-as-flow have no consistent meanings, in practice, across different research traditions. In part, those differences are due to different histories of those fields. In part, those differences reflect different experiences with ethical frameworks, such as the Common Rule mentioned earlier. With respect to making productive uses of data, some of these differences and complexities can be bridged via computational techniques.[99] Others can addressed by research strategies that implement "de-composability" ideas, by building research products that interoperate in modular ways with research products from other fields, like Lego bricks.[100] But commons

90   Strandburg, Frischmann and Madison (n 54).

91   Simon (n 72).

92   Donald E Stokes, *Pasteur's Quadrant: Basic Science and Technological Innovation* (Brookings Institution Press 2011).

93   C Jung and others, 'Optimization of Data Life Cycles' (2014) 513 *Journal of Physics: Conference Series* 032047.

94   Goodman (n 88).

95   Frischmann, *Infrastructure* (n 50).

96   A McKenna and others, 'The Genome Analysis Toolkit: A MapReduce Framework for Analyzing next-Generation DNA Sequencing Data' (2010) 20 *Genome Research* 1297.

97   Marcel Boumans and Sabina Leonelli, 'From Dirty Data to Tidy Facts: Clustering Practices in Plant Phenomics and Business Cycle Analysis' in Sabina Leonelli and Niccolo Tempini (eds), *Data Journeys in the Sciences* (Springer 2020) https://ore.exeter.ac.uk/repository/handle/10871/40283 accessed 7 February 2020.

98   Tony Hey, Stewart Tansley and Kristin Tolle (eds), *The Fourth Paradigm: Data-Intensive Scientific Discovery* (Microsoft Research 2009).

99   Paul R Cohen, 'DARPA's Big Mechanism Program' (2015) 12 *Physical Biology* 045008.

100   David Singh Grewal, 'Before Peer Production: Infrastructure Gaps and the Architecture of Openness in Synthetic Biology' (2017) 20 *Stanford Technolo-*

governance strategies based on flexible understandings of the natures of research "things" provide an important set of tools, bringing these approaches together via a systems perspective.[101]

A second example comes from within the law, from intellectual property law and its treatment of data. Here, the problem is that treating a data resource as data-as-form or as data-as-flow in one IP system may push actors to change their characterization of resources with respect to a different system. Recently, the US Supreme Court invalidated patents on genetic sequences isolated from human genes, in *Association for Molecular Pathology v. Myriad Genetics*, Inc.[102] That ruling undercut the power of the patentee, Myriad Genetics, to build a commercial business around genetic testing based on identifying those sequences in individuals. Those who supported invalidation and advocated for eliminating patent coverage of genetic sequences cheered. This appeared to be a win for research science, for the concept of knowledge flow, and to many, for better clinical health outcomes and public health. Yet it appears that Myriad has adjusted its business strategy, applying non-patent strategies to enhance the exclusivity of the pools of research data that were used to develop the patented inventions.[103] What law seems to provide in one legal domain (data-as-flow), it seems to take away in another, at least in part (data-as-form). Similar conflicts now exists with respect to public sector uses of DNA data in criminal proceedings, on the one hand, and trade secrecy law, on the other hand,[104] and between public health objectives and efforts to protect patient privacy by granting property rights in personal data to individual patients.[105] An ecological or systems approach may not solve these specific problems, but it would allow policymakers to anticipate them more clearly.[106]

It should be emphasized that thing-ness or resource forms, whether given, designed, or constructed by law or otherwise, should not be viewed as necessarily hostile to efforts to promote data openness and data sharing. So long as the character and attributes of a knowledge resource are matters of design, including legal reinforcement or disruption of thing-ness, then the design of resources can be tailored appropriately to relevant governance goals. Building a data repository of shared scientific data, for example, typically requires coordination and collaboration as to technical matters (can one dataset be combined or coordinated with another dataset as matters of code?), as to legal matters (are enabling or disabling contracts, licenses, covenants, and/or laws present?), and as to social, cultural, and economic matters (do libraries and archivists and research scientists and institutional administrators each understand, appreciate, and respect how field-specific expertise and other resources are needed to ensure the utility and stability of the repository?).[107]

In conversations that embody those challenges, data-as-flow can be

a virtue (because giant data repositories can support streams of new, fantastic research), but it can also be a vice (because contributions of different fields and different resources may be difficult to separately identify and manage, in practice). With things as with groups, no panacea exists, that is, no "one size fits all" solution. Data governance counsels taking an adaptable stance on data-as-form and data-as-flow questions, rather than a rigid or ontological one.

6.   Looking Ahead

This article offers a conceptual toolkit for data governance that centers on two big themes: groups and things. Those can be combined in various ways as part of developing approaches to governance data collection, production, storage, stewardship, and use. Knowledge commons is proposed as a significant overarching framework for using these tools in developing data sharing strategies, but the tools are also relevant to understanding market-based or state-based institutional governance. As a conceptual approach, the pair of tools comes with few necessary payoffs or implications. For example, stereotypical lessons such as "define resources with clarity" or "determine boundaries regarding access and use with specificity" may have grounding in research on natural resources by Ostrom and others, but perspectives on knowledge and information resources teaches that different guidance may apply in those contexts, or some of them.[108] The path forward lies as much in imaginative use of the concepts described here as in specific rules for specific problems. Four possible imaginative uses follow.

6.1   Examine Social Groups and Resources in Systems

Neither data governance nor knowledge commons should be implemented in a single way across all fields and domains. Large-scale initiatives to promote openness in research science, AI systems, urban planning, public administration and law, environmental regulation, and public health face the difficult but critical challenge of inventorying, understanding, and analyzing the technical, social and cultural, and legal attributes of polycentric ecologies. Data governance implies that collaboration strategies should be built out of those details.

That implication applies to private collaboratives such as the Open Data Initiative supported by Microsoft and other technology companies,[109] and to individuals and enterprises advancing the Panton Principles, calling for open data in science.[110] It applies to global NGOs focused on forward-looking uses of data such as AI for Good,[111] and private counterparts such as AI Commons[112] and Open AI.[113] It applies to governments. It applies to individual firms, to universities and research organizations, and even to individual policymakers, researchers, data scientists, and archivists.

Relatedly, too much emphasis in developing effective and appropriate data governance may be put on traditional distinctions between public and private enterprises and public and private goods. Similarly, too much emphasis may be put on identifying and reinforcing distinctions between data and algorithms. Last, too much emphasis

gy Law Review 143.

[101] Robert Cook-Deegan and Tom Dedeurwaerdere, 'The Science Commons in Life Science Research: Structure, Function, and Value of Access to Genetic Diversity' (2006) 58 *International Social Science Journal* 299.

[102] 569 U.S. 576 (2013).

[103] John M Conley, Robert Cook-Deegan and Gabriel Lázaro-Muñoz, 'Myriad after Myriad: The Proprietary Data Dilemma' (2014) 15 *North Carolina Journal of Law & Technology* 597.

[104] Sonia Katyal, 'The Paradox of Source Code Secrecy' (2019) 104 *Cornell Law Review* 1183.

[105] Jorge L Contreras, 'The False Promise of Health Data Ownership' (2019) 94 *New York University Law Review* 624.

[106] Helen Fay Nissenbaum, *Privacy in Context: Technology, Policy, and the Integrity of Social Life* (Stanford Law Books 2010).

[107] Max von Grafenstein, Alina Wernick and Christopher Olk, 'Data Governance: Enhancing Innovation and Protecting Against Its Risks' (2019) 54 Intereconomics 228.

[108] Madison, 'IP Things as Boundary Objects' (n 77).

[109] https://www.microsoft.com/en-us/open-data-initiative.

[110] https://pantonprinciples.org/index.html.

[111] AI for Good is a United Nations platform for dialogue on future uses of artificial intelligence https://aiforgood.itu.int/.

[112] AI Commons is a non-profit organization collecting diverse contributions to ensure that the benefits of AI systems are broadly distributed. https://aicommons.com/.

[113] OpenAI is a private enterprise whose mission is to ensure that AI systems benefit all of humanity. https://openai.com/.

may be placed on the idea of data as an infrastructural resource and on data infrastructures. None of those distinctions are unimportant. How they are advanced, or modified, are questions for governance discussions.

6.2   Build Pragmatic Models of Policy Problems

Data are sometimes viewed optimistically, as enabling spillover individual and social benefits, and sometimes skeptically, as constraining individuals or imposing harms. An institutional governance framework supplies a useful method of integrating these different and sometimes disparate perspectives into a pragmatic, systems-based matrix.

Efforts to "regulate" data production and use via public/private matrixes or on a field-by-field basis have often proved to be inadequate or inflexible, because regulators, policymakers, and scholars have too often tried to squeeze something that "looks and feels" like an intellectual resource into the IP categories that were constructed over the course of the twentieth century for other intellectual resources: copyright, patent, trade secrets and confidential information, and related fields such as antitrust and unfair competition, and privacy.

Positive law is thus seen in part as providing ways of solving social dilemmas regarding shared resources such as data, by encouraging collaboration via supplying state subsidies for infrastructure; creating safe harbors for commercial collaboration and exemptions from unfair competition and antitrust charges; exempting information from exclusionary IP regimes; offering convening and facilitation services; and in other ways.[114] Positive law is also sometimes seen as impeding collaboration, creating social dilemmas rather than solving them. The idea of the anti-commons, in which a social space is characterized by too many separate property claims recognized by law, is one suggestive example.[115] An approach that organizes data regulation by traditional legal field struggles to reconcile those perspectives.

A promising model for integrating them and others, using a pragmatic approach based on a governance rubric, is the work of the political scientist Martha Finnemore and the legal scholar Duncan Hollis on constructing "cybernorms" for global cybersecurity governance.[116] They argue that managing global cybersecurity data is a systemic and ecological problem; that it does not fit standard policy-specific boxes for diagnoses or solutions; and that polycentric, group-based strategies are most likely to be effective on grounds of legitimacy and adaptability.

6.3   Expect Change, and Borrow From Experience

A pragmatic approach to data governance makes explicit that governance mechanisms must be adaptable, and they must be adaptable at different scales (small to large, slow to fast, local to global, existing to novel) and relative to different resources (human capabilities, social and institutional capabilities, and technological capabilities).

That emphasis on adaptability brings out a possibly surprising feature of governance, and in particular data governance, that focuses on social groups and on things: its receptivity to established governance

mechanisms, even those that long pre-date the rise of Big Data, the internet era, or even twentieth century technology. Contemporary IP researchers have acquired a recent interest in informal, norm-governed innovation communities,[117] where formal systems of IP rights as such seem to contribute little or not at all to developing bodies of novel and creative work.

That interest in collective creativity can be traced back to the earliest days of research science, in the Republic of Letters and the early Enlightenment in England, Scotland, and continental Europe. Communities of scientific researchers formed face to face and correspondence networks, eventually becoming formalized in salons, scientific societies, and journals. This was not the practice of formal peer review. It was, instead, a polycentric network of social groups, regulating itself and the contents of their contributions via a complex system of social norms.[118] That centuries-old style of knowledge commons governance has been durable, adaptable, and effective. It may be relevant today.

6.4   Build Assessment Techniques

Perhaps the most difficult challenge to confront in designing and analyzing data governance is the question of assessment. Institutional design is significantly a question of comparative analysis. By what measure is one governance institution preferred to another?

Political theory, economic theory, and social theory have no shortage of answers. Social welfare analysis gives us attention to outputs (utility, including spillovers) and to inputs (human capabilities and capacities). Social choice theory asks us to assess the character of processes of collective choice regarding institutional arrangements. Should institutions aggregate or otherwise accurate reflect the preferences of their participants? Political philosophy directs us to ask questions about legitimacy, transparency, accountability, and protection of primary values of individual human autonomy, including powers of self-determination regarding participation in the polity.[119]

For example, a data governance community that sustains itself in coordination with the state differs from a nominally open community that proceeds only by relying on state-sanctioned legal instruments. Modern scientific research has the former character, given the abundant direct support and tax benefits offered to scientific research institutions and researchers themselves. Users of the Creative Commons licensing tool likely have the second character; mere use of a Creative Commons license, taken alone, does not enroll the user in a collective or community of any sort, and the license instrument itself is a salient and near cousin of proprietary licenses.[120] A group that manages an "open" resource, such as data, entirely via legal instruments, is apt to encounter incompatibility problems. Not every open data license defines "open" the same way.[121]

[114] Jorge L Contreras, 'Leviathan in the Commons: Biomedical Data and the State' in Katherine J Strandburg, Brett M Frischmann and Michael J Madison (eds), *Governing Medical Knowledge Commons* (Cambridge University Press 2017).

[115] Michael A Heller, 'The Tragedy of the Anticommons: Property in the Transition from Marx to Markets' (1998) 111 *Harvard Law Review* 621.

[116] Martha Finnemore and Duncan B Hollis, 'Constructing Norms for Global Cybersecurity' (2016) 110 *American Journal of International Law* 425.

[117] Kate Darling and Aaron Perzanowski (eds), *Creativity without Law: Challenging the Assumptions of Intellectual Property* (NYU Press 2017).

[118] Michael J Madison, 'The Republic of Letters and the Origins of Scientific Knowledge Commons' in Madelyn Sanfilippo, Katherine J Strandburg and Brett M Frischmann (eds), *Governing Privacy Commons* (Cambridge University Press forthcoming).

[119] Hanoch Dagan and Michael A Heller, 'The Liberal Commons' (2001) 110 *Yale Law Journal* 549.

[120] Niva Elkin-Koren, 'Creative Commons: A Skeptical View of a Worthy Pursuit' in Lucie Guibault and P Bernt Hugenholtz (eds), *Future of the Public Domain* (Kluwer Law International 2006).

[121] Alexandra Giannopoulou, 'Understanding Open Data Regulation: An Analysis of the Licensing Landscape' in Bastiaan van Loenen, Glenn Vancauwenberghe and Joep Crompvoets (eds), *Open Data Exposed*, vol 30 (TMC Asser Press 2018).

Measures of experience on the ground matter. Does knowledge commons governance work? Is governance durable and sustainable across time (generations) and space (relevant state and other organizational boundaries and borders)? Does practice align with relevant ideology, including relevant rhetorics, enhancing not only its descriptive legitimacy (acceptability to the community and to society) but also its normative claims?

The adaptability, flexibility, and even fuzziness of commons governance in information and data settings makes assessment even trickier. Stipulating that data-as-form and data-as-flow are key governance attributes, that data may exist in multiple interpreted forms and flows simultaneously, and that resources and groups are often engaged in projects of producing and re-producing one another, complicates classic governance distinctions between individuals and collectives, people and things, and subjects and objects.

7.   Conclusion

In almost all contexts of interest for data governance purposes, data are likely to be shared. When, how, and why to share data are governance topics. This article has argued that the fundamental yet nonetheless pragmatic governance question for data is understanding different implications of seeing data-as-form and data-as-flow.

This is a conceptual argument. It is undoubtedly true that where law meets technology, whether on economic grounds or social and cultural terms, rules matter. Positive law matters, along with systems of social norms, customs, and conventions. Rights and interests matter, and their integration into regulatory frameworks matters, too. Nonetheless, the article recommends beginning not with the rules but with questions of institutional design, motivated by key concepts. A well-grounded domain of research exists focusing on shared knowledge, information, and data as objects and subjects of institutional governance. That domain is knowledge commons. Knowledge commons analysis argues for identifying and describing relevant social groups in which governance frameworks may be embedded, and for identifying and describing relevant resources, or things, whose form and flow will contribute substantially to the welfare effects of the relevant data governance systems. Those are tools for data governance.

This perspective takes an ecological or systems approach to regulatory questions, an approach in which market exclusivities and state mandates do not provide the standard two-part regulatory framing. Knowledge commons governance, in which data and information resources are shared according to governance rules tied to identified social and institutional collectives, provides a substantial third storehouse of data governance solutions.