2020

Data Governance and the Emerging University

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

Data governance and the emerging university

Draft of 5 November 2019

Michael J Madison

I. Introduction

As data become more central to university research practice, data governance policies implicate questions that are more fundamental than those supplied by framings grounded in existing technology transfer programs. Should data be "open" or "closed"? Are data "basic" knowledge or "applied" knowledge? Neither question suits data and data practices themselves. Governing data suggests re-thinking the character of universities as research institutions.

What should universities do with research data collected and generated by their researchers, and why? The question has both positive and normative attributes. Positively, the questions concern the complexities of data as an information resource in a specific setting. Normatively, the questions concern a novel way of illuminating continuing challenges associated with governance of the university itself, as a knowledge-based enterprise. "What should be done with data?" is a mode of asking "what are the purposes of universities?" This Chapter weaves these two themes together. Combined, they represent both a forward-looking inquiry into a still-critical institution and a case of addressing broad and abstract social questions via understanding their institutional contexts.

Positively and pragmatically, any modern research university has at hand vast quantities of data and datasets, much of them collected and generated by the university's own faculty and students and stored locally, some of them accessed by the university's researchers but stored and maintained elsewhere. Data constitute critical research resources across multiple domains of inquiry and practice, rather than niche or specialized tools. Recognizing their cross-cutting significance, in Europe, in the United Kingdom, and increasingly in the United States and elsewhere, universities are preparing informal guidance and formal policies to assist researchers and others in storing, curating, securing, and sharing and distributing research data.[1] Funders, regulators, and scholarly publishers may expect universities to share data and may also expect universities to secure data. Researchers themselves often want both to share data and to control it. Amid a diverse university environment, what should data governance policies consist of?

Normatively, those pragmatic concerns overlap with advocacy among many researchers and information policy professionals of principles animating the

---

[1] *See*, *e.g.*, Martin Donnelly, *Five Steps to Developing a Research Data Policy*, DIG. CURATION CTR. (Jan. 2014), http://www.dcc.ac.uk/resources/policy-and-legal/five-steps-developing-research-data-policy/five-steps-developing-research.

ideas of Open Science and Open Data.[2] That advocacy takes some form of the following arguments. The enormous volumes of data that underlie contemporary research constitute a massive knowledge resource that ought to be openly accessible to researchers everywhere. It is argued that data constitute knowledge. To make knowledge useful, in various modern senses, knowledge ought to be open, and knowledge ought to be shared.[3] In the abstract, those propositions are nearly incontrovertible, but advancing and implementing them, and addressing complementary and sometimes opposing goals, means approaching them institutionally, in programs and practices rather purely as concepts. In what specific respects should data governance embody or reflect aspirations toward "openness"?

This Chapter sets those questions in the broader contexts of the university as an institution and of data as university-based resource. Highlighting that context implicates intersections and overlaps between emerging practices and goals regarding research data in universities, on the one hand, and established practices and expectations regarding intellectual property ("IP") in the university, particularly university-based inventing, patenting, and technology transfer, on the other hand. Much as the Open Science movement aims to encourage researchers, administrators, funders, publishers, and regulators to establish institutional frameworks to promote data sharing, at the same time existing institutions of university-based IP production may push, instead, toward embedding university research data in closed, proprietary frameworks and practices.

One purpose of the Chapter is to anticipate and describe that possible conflict using the relatively well-trod pathways that distinguish between open and proprietary frameworks in addressing how to manage research and scholarship in the university, as related modes of knowledge production. Those pathways include conceptual divides between the production and dissemination of basic or generalized knowledge, on the one hand, and useful or specific applications of knowledge, on the other hand. Those conceptual divides are traced institutionally via university-based forms and practices of IP law, namely patents and copyrights. Basic or generalized knowledge is documented and disseminated via scholarship. Scholarly works are governed by copyright law and are subject to dual normative imperatives: broad distribution and no expectation of reward or return to the researcher. Useful or specific knowledge is disseminated via being encoded into inventions, governed by patent law, and their dissemination is linked to rewards to both inventors and their university employers via technology transfer systems.

---

[2] *See, e.g.*, *Open Science by Design: Realizing a Vision for 21st Century Research*, NAT'L ACADS. OF SCI., ENG'G, AND MED. (2018); *Final Report – Science as an Open Enterprise*, ROYAL SOC'Y (2012), https://royalsociety.org/topics-policy/projects/science-public-enterprise/report/; *OECD Principles and Guidelines for Access to Research Data from Public Funding*, OECD (2007), http://www.oecd.org/science/sci-tech/38500813.pdf.

[3] *See, e.g.*, JOEL MOKYR, THE GIFTS OF ATHENA: HISTORICAL ORIGINS OF THE KNOWLEDGE ECONOMY (2004).

In short, "open" knowledge production corresponds roughly to university institutions associated with scholarly production and to copyright; "closed" knowledge production corresponds roughly to university institutions associated with inventions and patent.

In this scheme, where does data fit? The Chapter argues that like inventions (in part) and copyright works (in part), data and datasets constitute both modes of useful knowledge and of basic or general knowledge. Shoehorning data into patent-like technology transfer practices risks closing off access to and use of important, basic, research resources. But the alternative approach, given the either/or "open v. closed" premise, is unworkable for different reasons. Assigning data rhetorically and practically to the more open institutions and practices associated with copyright and scholarship risks harm to data-intensive research itself, by imposing a "thing"-like character on data governance when research practices often assume or rely on data's "flow"-like character.

A second purpose of the Chapter is to introduce data as a perspective on the evolving character of the research university itself, highlighting conflicts between openness and proprietary claims that build on a less common perspective. Universities today have many purposes, but central to almost all descriptions of the university is the production, collection, curation, and distribution of knowledge.[4] Traditionally and conventionally, the university exercised those functions through teaching, collecting physical specimens, and engaging in research and publishing scholarship. In other words, for centuries universities generated and shared knowledge through defined pathways of experience: via interactions among humans, via interactions between humans and objects, and via text.[5] Data-based research constitutes each of these, in one sense, and none, in a different sense.

As the university continues to evolve, the new salience of data in the research enterprise illustrates some central conflicts over what the university is and what it should be. Perhaps, at its core, the university is still a knowledge institution. If the university is now a data-oriented institution, as it now appears to be, then one might conclude that data are knowledge. That's a functional claim rather than an ontological one. As a basis for governance, it falls short, because for the reasons suggested earlier, the "open v. closed?" framing of university-based knowledge production suits data poorly. Perhaps the foundational assumption of modern research universities, that they should distinguish between basic or general knowledge and useful or specific knowledge, should be re-thought. Perhaps, instead, the university is no longer primarily a knowledge institution;

---

[4] *See* JAROSLAV PELIKAN, THE IDEA OF THE UNIVERSITY: A REEXAMINATION (1992).
[5] *See*, *e.g.*, Michael J. Madison, Brett M. Frischmann, & Katherine J. Strandburg, *The University as Constructed Cultural Commons*, 30 WASH. U. J. LAW & POL'Y 365, 381-86 (2009).

data themselves supply the university's new central organizing principle, and entirely new governance principles should be developed.

Organizationally, the Chapter begins by laying out briefly its chief pragmatic concern: the rise of data-intensive research and the role of data and datasets in the modern university. The problem to be described, as it were, is how the university ought to manage those resources for maximum social benefit and minimal social and private harm. The Chapter takes a pragmatic approach rather than a first principles approach. General questions about the virtues and drawbacks of data collection and of data sharing and data security are largely left for others to explore. Here, questions relate primarily and more concretely to what should be done as data governance in a specific institutional setting, and how, and with what implications.

Having described the character of modern data-intensive research, the Chapter turns to a brief account of the university itself, tracing themes of openness and the character of university-based knowledge through history, to the present.

The Chapter then combines themes offered by the review of data-based research and the history of universities. Claims of openness attached to emerging data-based research practices in the 21st century university stand in contrast to the proprietary institutions of patent law and technology transfer carried over from the late 20th century university. But that statement of "open v. closed?" affirms a case of institutional continuity as much as it offers a case of contrast. The university has wrestled for centuries with conceptual and practical demands that applied knowledge research practices be accommodated as complements to basic knowledge research practices; the modern research university is a dynamic amalgam of open and closed governance.[6] The 20th century university has been criticized on the ground that its open, knowledge-driven mission is unduly influenced by demands that knowledge be controlled and proprietary. Emerging interest in data governance may push toward a renewed focus on openness. But trying to situate data in that framing may do more harm than good. Instead, the Chapter offers the concept of the data-intensive university, a fundamental reframing of institutional purpose that permits asking questions about data governance that neither assume nor reject "open v. closed?" (or "basic v. applied?") as foundational premises but instead place governance questions firmly in new institutional context.

II. The meanings and applications of data-intensive research

Data challenge conventional distinctions between basic or general knowledge and applied or useful knowledge. Data have long been understood to consist of important inputs into research programs, as well as research tools or

---

[6] DONALD E. STOKES, PASTEUR'S QUADRANT: BASIC SCIENCE AND TECHNOLOGICAL INNOVATION 26-57 (1997).

instruments. Today, data may also constitute important research outputs. That plural character challenges the application of standard governance frameworks grounded in IP and other information law.

Discussions of data and data governance should begin with some history and some cultural and legal context. The rise of research data and the problem of knowledge governance concerning data are woven from threads of different colors. This Part addresses both established and novel attributes of data in research, and it outlines relevant frameworks in law.

A. Data: what's old, what's new

Scholars and scholarship begin with data. They always have. Historians organize and interpret events in the world. Geologists organize and interpret attributes of the earth. Biologists and psychologists organize and interpret attributes of life forms. What has changed in recent decades is the power of computing technologies. Creating and interpreting massive digital datasets is now within the reach of both specialized fields and disciplines that are organized around those computational problems and also pre-existing fields and disciplines that have begun to adopt and explore computational techniques. Gathering, processing, and interpreting super-sized datasets is becoming the norm in many fields, not just a few.

"Data" should be defined, initially but broadly, as *evidence,* or generally, as "representations of phenomena in the world," natural, physical, material, and cultural, used in connection with research and scholarship.[7] That is a functional definition. Data are gathered, created, and used by researchers as informational inputs into interpretive processes, practices, and technologies that themselves are useful as means of generating and sharing semantically meaningful results. Data may be infrastructural resources, in the sense that they provide bases for multiple applications, interpretations, and other uses in complex systems.[8] As inputs, data may be regarded as kinds of evidentiary raw material. As infrastructural resources, data may be regarded as kinds of multi-purpose research tools, or instruments.

By convention, we often treat "data" as a foundational resource and then call the interpreted results "information" or "knowledge," implying a typically linear or hierarchical relationship. That conventional linearity or hierarchy has attributes of prestige as well as function. In both senses, the convention has an

---

[7] *See* CHRISTINE L. BORGMAN, BIG DATA, LITTLE DATA, NO DATA: SCHOLARSHIP IN THE NETWORKED WORLD 17-29 (2015).

[8] *See* Paul N. Edwards, et al., *Knowledge Infrastructures: Intellectual Frameworks and Research Challenges (*Report of a workshop sponsored by the National Science Foundation and the Sloan Foundation, University of Michigan School of Information), May 2012, at 25-28; BRETT M. FRISCHMANN, INFRASTRUCTURE: THE SOCIAL VALUE OF SHARED Resources 225-26 (2012).

ancient lineage. At least beginning in the 17^{th} century, and perhaps as long ago as Greek antiquity, the fruits of observation, investigation, and study were divided between basic, or general, or eternal principles governing the world, considered to be knowledge, the province of philosophers; and applied or useful information, the province of mere engineers and technologists. "Data" and the singular "datum" emerged during the 17^{th} century as part of the syntax of Baconian science, but its plural uses gave it life in both basic science and applied technology domains. The phrase "data processing," now somewhat out of fashion, captured a mid-20^{th} century sense in which "data" were considered to be "raw material" for higher-order industrial analysis.[9]

Even in traditional framings, there has rarely been a hard or fast divide in practice between "data" and "knowledge." Rather than thinking of data as parts of linear pathways leading to knowledge, both historians of science and scholars of information believe that it is more appropriate to understand data as parts of overlapping processes of evidence collection and interpretation. Even if data are in part *form* that contribute sequentially to the production of information and knowledge, data are equal part *flow*, in which evidence of phenomena lead to interpretation and possible use, perhaps being documented and disseminated in various *forms*, and those forms of knowledge and information in use provide further evidence for further interpretation, and so on.

For example, in many fields, sociologists of science teach that a sense of data as solely or primarily "raw material" will recede from explicit consideration via long processes of embedding observational details in domain-specific concepts and categories. A "wave form" was "data" once, to physicists, but today it may simply be a wave form, whose meaning and function are self-evident within the field. Data cease being things or objects in themselves. They are codified—black boxed, sociologists might say—as part of disciplinary practices.[10] Data are not "data"; they are research. Datasets and data collection as such are increasingly viewed in some scholarly domains as credit-worthy research and scholarly activities in their own rights. Some both inside and outside the university setting argue that production and publication of a complex dataset ought to be treated as a meritorious knowledge producing activity, eligible for consideration in recruitment, retention, and promotion decisions for students and researchers.[11] Interventions in those debates vary from field to field, as one would expect. Not everyone agrees that publishing a novel dataset should be treated by a field as equivalent to publishing a meritorious journal article.

---

[9] *See, e.g.*, JOANNE YATES, CONTROL THROUGH COMMUNICATION: THE RISE OF SYSTEM IN AMERICAN MANAGEMENT (1989).
[10] *See* BRUNO LATOUR, SCIENCE IN ACTION: HOW TO FOLLOW SCIENTISTS AND ENGINEERS THROUGH SOCIETY 2-3 (1987).
[11] Borgman, *supra* note 7, at 47-53.

Newish phrases such as "data curation" and "data carpentry,"[12] and practices of data visualization,[13] express related sentiments. Data are not simply "out there" waiting to be found, but have to be assembled and shaped, as cultural phenomena. Data and datasets are interpretation as much as representation. Judith Donath offers the phrase "data portraiture" as an alternative to "data visualization" to call attention to the possibility that "visualization" suggests a practice that might be regarded, inaccurately, as free from professional judgment, or that data, like human identity, does not express itself. Efforts to interpret and represent data (themselves, representing some other phenomena) have foundations in the arts as well as in more conventionally technical fields.[14] Data and datasets may be handmade. Data and datasets may be generated by algorithms. In either case, they nearly always represent disciplined choices about how material in the world is gathered, combined, made salient (or not), stored, and communicated. They are themselves modes of basic or generalized knowledge, meaning that data constitute research results and products generated either by domain-based researchers or by specialized researchers in the field increasingly known as data science.

The proposition that data are simultaneously input, tool, and generalized knowledge is hinted at by the related idea of data as *flow*. That concept has now been institutionalized by scholars via the information data lifecycle, which rejects the idea of data as solely "raw material," "stuff," or fixed "object," leading to interpreted knowledge, in favor of an ecological approach.[15] Data are representations not of a particular state of the world (as a scientific researcher might consider them), but instead representations of intersections among evolutionary processes that define the relevant natural or social world and the ecological processes by which researchers identify, create, curate, and disseminate data. Some designers of health care information technology systems build on that premise by promoting "data liquidity": ensuring that the right data is accessible to the right person, at the right time.[16]

Exploring data as flow as well as form takes on new significance in the research enterprise by virtue of the rise of data-intensive research, sometimes

---

[12] *See* Tracy K. Teal, et al., *Data Carpentry: Workshops to Increase Data Literacy for Researchers*, 10 INT'L J. DIGITAL CURATION 135 (2015), doi:10.2218/ijdc.v10i1.351.

[13] *See* EDWARD R. TUFTE, VISUAL EXPLANATIONS: IMAGES AND QUANTITIES, EVIDENCE AND NARRATIVE (1997); EDWARD R. TUFTE, ENVISIONING INFORMATION (1990); EDWARD R. TUFTE, THE VISUAL DISPLAY OF QUANTITATIVE INFORMATION (1983).

[14] *See* JUDITH DONATH, THE SOCIAL MACHINE: DESIGNS FOR LIVING ONLINE 209 (2015).

[15] *See*, *e.g.*, Alex Ball, Review of Data Management Lifecycle Models, REDm-MED Project Document (U. Bath, 2012), doi: redm1rep120110ab10.

[16] *See* Paul K. Courtney, Data Liquidity in Health Information Systems, 17 CANCER J. 219 (2012), doi:10.1097/PPO.0b013e3182270c83.

referred to as data-intensive science or "Big Data."[17] Whatever the label, the stakes of data governance are greater and more explicit than ever before.

Data-intensive science has its roots in the fact that beginning roughly twenty-five years ago, a handful of specialized scholarly disciplines began to grow up specifically to deal with opportunities to explore super-large digital datasets, particularly in particle physics, astronomy, and astrophysics, on the one hand, and genomics, on the other hand. In each case but in different respects, customized information technologies were developed to process exceptionally large collections of observations of the world, at super large scales (astronomy and astrophysics) or at super small scales (particle physics and genomics). In difference respects, researchers linked observational capabilities, high speed communications and computer networks, and storage and image processing technologies, sometimes including high capacity systems for converting information resources from analog (material) to digital (virtual) forms.

Notably, the datasets themselves were not the province of university-based researchers operating alone. The scale of the research enterprise meant that complex partnerships to produce and manage datasets developed among academic researchers, researchers in other settings, philanthropies, national governments, and publishers.[18] Calling additional attention to contributions from industry, Caroline Wagner calls this cluster of developments "the new invisible college," describing science and technology "as an emergent networked system rather than as a national asset."[19] In the new invisible college, the idea of data as *flow* takes on a set of concrete institutional forms.

Moreover, abstracted from these specific domains, powerful new computational resources and their descendants—technologies for sensing and observing, information processing and storage capabilities, network communications facilities, and related interpretive techniques—have been introduced gradually to and accepted in other research domains, such as the so-called digital humanities. Ever-increasing swaths of the university now fall under the general heading of "data-intensive" science, discovery, research, and collaboration,"[20] because they collect or generate data or create complex models for analyzing and presenting data. Researchers of many stripes are now data scientists in both name and function.

---

[17] *See, e.g.*, *Data-Driven Innovation: Big Data for Growth and Wellbeing*, OECD (2015), www.oecd.org/sti/data-driven-innovation-9789264229358-en.htm (last visited Mar. 28, 2019).

[18] *See Open Science by Design*, *supra* note 2, at 78-100.

[19] CAROLINE S. WAGNER, THE NEW INVISIBLE COLLEGE: SCIENCE FOR DEVELOPMENT 9 (2009).

[20] *See* THE FOURTH PARADIGM: DATA-INTENSIVE SCIENTIFIC DISCOVERY (Tony Hey, Stewart Tansley, & Kristin Tolle eds., 2009); Borgman, *supra* note 7, at 31.

Given its technological depth and breadth, the data lifecycle increasingly needs and gets a lot of higher status—that is, research-based—technological and financial support. Research data management now has its own acronym: RDM.[21] Commercial interests may occupy critical positions in the data lifecycle. A key source of bioinformatics research data in medicine and public health comes from clinical medical practice. Universities with medical schools and allied medical providers as partners may rely on those institutions as sources of research data. Providers' and insurers interests in expense and revenue streams, and in commercial opportunities relevant to clinical practice, may collide with researchers' and scholars' interests in that data. Universities may structure technology transfer practice in ways that prioritize claimed proprietary interests in the data over research interests in sharing the data.[22]

In sum, data have now fully lost their conventional resonance as so-called "stuff that scholars interpret" and have acquired a new, broad, and distinctively modern technological resonance as both resources for interpretation and resources that result from interpretation, and in which a great deal of specialized time and equipment are invested. In a sense, data remain research tools. But as research tools they are critical knowledge resources throughout the research enterprise, not only in laboratories, and they are important research products as well as research infrastructures. Data, datasets, and related computer software and hardware are no longer the special province of a handful of researchers. They are the keys to the entire university. Its research data resources dwarf its potentially patentable resources, in technical scale and social significance.[23]

B. Data and the law: conflicts with IP premises

Lawyers, legal scholars, administrators, and policymakers focus on law, regulation, and governance of data and datasets. Governance, as that term is used in this Chapter, means opportunities for and limitations on data collection, production, storage, and use by the university and its researchers, including both formal rulesets imposed from the outside (the state, funders) and the inside (the rules and policies of the university), plus informal norms and expectations emerging from various sources (disciplinary expectations and practices both inside and outside the university). Data governance, which might be regarded as a species of knowledge governance or information governance,[24] embraces this multiplicity

---

[21] *See*, *e.g.*, *Research Data Management,* ONLINE COMPUTER LIBRARY CENTER ("OCLC"), https://www.oclc.org/research/themes/research-collections/rdm.html (last visited Mar. 28, 2019).

[22] *See* Jacob S. Sherkow & Jorge L. Contreras, *IP, Surrogate Licensing, and Precision Medicine*, 7(2) IP THEORY 1 (2018).

[23] *See* Borgman, *supra* note 7, at 125-202 (describing data-related research practices in the social sciences and humanities).

[24] *See* Michael J. Madison, *Reconstructing the Software License*, 35 LOY. U. CHI. L.J. 275 (2003).

of interests, as a complex, changing system that answers the questions: who uses and should use data, and how?

One critical point is how data governance is colored heavily by IP concepts, in general and specifically within the university. Concretely, that means patent law and copyright law. The IP origin story matters here. In society and commerce as a whole, that it, from perspectives external to university and scholarly practice, many lawyers, policymakers, and commercial interests have long advanced the conceptual argument that "creative" and "innovative" intellectual production, or modalities of knowledge defined over time by relevant bodies of law, are reducible to conceptual things: copyrightable works and patentable inventions. In the language of the law, those things are protected from improper appropriation and are tradeable in markets because of their expected economic value. IP law "thing-ifies" or codifies intellectual production precisely in order to stimulate the production of more, different, or better intellectual "things."[25] Sometimes that "thing-ification" instinct leads to good results in both large and small senses, and sometimes it leads to poor ones,

How do those IP concepts, and the thing-ification instinct, intersect with data and datasets? The logic is straightforward. Generating or collecting data is a mode of intellectual production, because data is either a mode of information production or a step in knowledge production, or both. According to one typical line of reasoning, data are or should be "thing-ified" in the eyes of the law, from an IP standpoint, if and to the extent that "thing-ification" would help society achieve what it wants: more data (intellectual production), or new data, better data, or more valuable data.[26] This would give intellectual "things" status akin to the property law protection usually given to manufactured objects. A different line would deny legal "thing" status to data as such, and would assign data to an intellectual public domain, if doing so would promote the production of more or better inventive or creative things *based on* that data. In an important sense, this would treat data as found, not made.

In most respects, IP law favors treating data and facts as un-owned, open, and freely shareable.[27] Property rights in data and datasets as such strike many people, particularly scientific researchers, as anathema to the belief that truths about the world are, as United States Supreme Court Justice Louis Brandeis once

---

[25] *See* Dan L. Burk, *The Role of Patent Law in Knowledge Codification*, 23 BERKELEY TECH. L.J. 1009 (2008); Michael J. Madison, *Law as Design: Objects, Concepts, and Digital Things*, 56 CASE W. RES. L. REV. 381 (2005).
[26] Within information science, discussion of information-as-thing has had a different character. *See* Michael Buckland, *Information as Thing*, 42 J. AM. SOC. INFO. SCI. 351 (1991).
[27] *See* Pamela Samuelson, *Mapping the Digital Public Domain: Threats and Opportunities*, 66 L. & CONTEMP. PROBS. 147 (2003).); Jerome H. Reichman & Pamela Samuelson, *Intellectual Property Rights in Data?,* 50 VAND. L. REV. 52 (1997).

wrote, "free as the air to common use."[28] Normatively, facts and data about the world are too useful and meaningful to too many people and for too many purposes, and their ontology tends too much toward "basic" and "found" rather than "applied" or "made," for proprietary claims to facts and data to be justified.

IP law takes that reasoning to more specific levels, holding to the premise that data both are and should be presumptively "open" but also offering some nuance and some basis for narrow claims that IP law might "protect" data under some circumstances. Data may appear to be "things" in form, but their thing-ness typically is linked to legal openness. Data are often regarded in law as facts about the world, including natural phenomena in the world. So long as that is the case, then data are not subject to IP "protection." In copyright, data and facts lack the creativity needed to justify legal coverage.[29] If, however, facts are interpreted or "created" in a copyright sense, because they are assembled or collected in a "creative" way (even a modestly creative way), then copyright may apply.[30] Claims of proprietary patent right relative to data and datasets are likely to be rejected on one or more statutory grounds. Datasets do not constitute "inventions" under Section 101 of the US Patent Act because they are mere "abstract ideas," or they may not be "useful" in the sense that the law requires. If, however, a data structure is "inventive," particularly in the context of computer programs, they may be patented.[31] In genomics, datasets may consist wholly or partly of genetic information. Data consisting of genetic sequences as such are usually considered to be unpatentable "products of nature," but synthesized genomic products may be patented.[32]

Framing data as open and unprotected in the IP context relies heavily on a conventional sense of facts as basic "raw material," matter that is foundational to interpreted knowledge products. Both that conventional sense and its embodiment in IP practice resist the earlier characterization of data as "flow" rather than "form" or "thing." That's one sense in which IP law is a poor match for data governance, even if IP law shares a "data openness" sensibility. There is a second sense in which IP law handles data governance poorly, in this case from a "proprietary data" sensibility. The conventional understanding of data as unownable "raw material" runs contrary to the idea, now common in modern scholarship, that facts may be

---

[28] Int'l News Serv. v. AP, 248 U.S. 215, 250 (1918) (Brandeis, J., dissenting); *see also* Yochai Benkler, *Free as the Air to Common Use: First Amendment Constraints on Enclosure of the Public Domain*, 74 NYU L. REV. 354 (1999).
[29] Feist Publ'ns, Inc. v. Rural Tel. Serv. Co., 499 U.S. 340 (1990).
[30]*See* Michael W. Carroll, *Sharing Research Data and IP Law: A Primer*, PLOS BIOLOGY 13(8): e1002235 (2015).; Justin Hughes, *Size Matters (Or Should) in Copyright Law*, 74 FORDHAM L. REV. 575 (2005).
[31] Enfish, LLC v. Microsoft Corp., 822 F.3d 1327 (Fed. Cir. 2016).
[32] Ass'n for Molecular Pathology v. Myriad Genetics, Inc., 569 U.S. 576 (2013).

interpreted knowledge products in their own right,[33] and to the instinct in the industrial sector that datasets may constitute commercially valuable, ownable products.

Some legal actors have been encouraged to figure out a better match for data governance than IP law, without lasting success in the structures of the law or in practice. First, data may be gathered and bundled in ways that make datasets amenable to treatment as assets that are subject to contract law, offering some form of legal protection against unauthorized data access or data use even to commodified collections of data as legal "things" —even while these "things" are excluded from IP coverage as a legal matter. That practice is increasingly accepted in industry and finance and in commercial domains, even if (and sometimes because) contract law offers a pragmatic yet fragile end-run around the public policy that dictates no IP coverage for data.[34] Second, in the context of biological product and drug development, statutory "data exclusivity" may attach to the results of testing new products independent of the patent status of the products themselves.[35] Third, in Europe, in 1996 the European Union adopted a Continent-wide Directive on protection of databases that was aimed at preventing "misappropriation" of data collections produced with meaningful investments of time, money, and/or expertise.[36] That instrument has had little practical effect. The Directive now faces possible revision.[37]

Despite this patchwork of formal governance instruments, in US law and elsewhere, the relative ineffectiveness of these strategies belies the fact that they reflect an instinct that is often common both to organizational life and to personal experience in the world, that data may be a thing. Not an unowned, open, "raw" thing as described above, but rather an asset—a manufactured thing, an object of possible economic, social, and/or cultural value, and a thing that is ripe for enclosure. That instinct co-exists with the instinct common in the research worlds that data is and ought to be regarded as part of a lifecycle, or flow.

The next Part situates these sensibilities and conflicts about data and data governance in related conflicts in the university.

---

[33] *See* MARY POOVEY, HISTORY OF THE MODERN FACT: PROBLEMS AND KNOWLEDGE IN THE SOURCES OF WEALTH AND SOCIETY (1998).

[34] Stacy-Ann Elvy, *Commodifying Consumer Data in the Era of the Internet of Things, Commodifying Consumer Data in the Era of the Internet of Things*, 59 B.C. L. REV. 423 (2018); Stacy-Ann Elvy, *Paying for Privacy and the Personal Data Economy*, 117 COLUM. L. REV.1369 (2017).

[35] *See* Erika Lietzan, *The Myths of Data Exclusivity*, 20 LEWIS & CLARK L. REV. 91 (2016).

[36] Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the Legal Protection of Databases, 1996 O.J. (L 77) 20.

[37] *See* European Commission, Staff Working Document and Executive summary on the Evaluation of the Directive 96/9/EC on the Legal Protection of Databases, Apr. 25, 2018, https://ec.europa.eu/digital-single-market/en/news/staff-working-document-and-executive-summary-evaluation-directive-969ec-legal-protection.

III. Data and the changing university

Contemporary university-based technology transfer institutions, grounded in research funding practices and IP regimes, encode a long-standing distinction in the university: between useful knowledge, which is controlled via the system, and basic knowledge, which is disseminated widely via the copyright system, as scholarship. In that framing, data may constitute both useful knowledge and basic knowledge. Yet data resist characterization by analogy to copyrightable scholarship or patentable inventions. Data governance requires a new framework, one that extends current conceptions of research in the university.

This Part introduces data and data governance as modes of knowledge production in the research university. In light of a foundational framing of the purposes of the university through history, it explores the inadequacy of the university's current parallel governance systems of open knowledge and closed or proprietary knowledge as applied to data. It considers the social and institutional risks of applying the current dual institutional structure too quickly to data governance questions. It suggests that the way forward includes reconsidering either what knowledge means, in the university setting, or what the university means, with respect to data and knowledge, or both.

A. An abbreviated history of the university

Long prior to and independent of concerns about data and datasets, the university has been characterized primarily by a commitment to knowledge, driven not by partisanship, bias, utility, or financial reward but instead by the desire for general understanding, and perhaps, by an interest in improving both human and natural conditions. In a strong sense, the university is still defined by the production of knowledge and by its dissemination throughout society.[38] But that definition is a puzzle, because knowledge production in the modern research university has a dual character. There is both research and scholarship, on the one hand, and also practices of invention and technology transfer, on the other hand. The modern university does not resolve that puzzle. Instead, it institutionalizes the duality via parallel practices of knowledge governance. Basic research, knowledge, and scholarship are meant to be driven by an ethos of curiosity, creativity, and experimentation, and their results are meant to be open and widely shared so that additional researchers can build on them in the same spirit. Applied research, knowledge, and invention are expected to be motivated by interests in solving specific technical or social problems, and their results are meant to be distributed via propertization and transfer from the university into market institutions. The institutional duality may be merged into an overarching commons governance

[38] See Alfonso Borrero Cabal, *The University as an Institution Today: Topics for Reflection*, UNESCO PUBL'G AND INT'L DEV. RESEARCH CTR. (1993).

framework, where commons refers to plural institutional practices of knowledge sharing.[39]

That duality blends with the university's dual institutional identity as a stable, changing entity and as a dynamic, adaptable one. In the 21st century, it is both imperative, in a sense, and inappropriate, in a different sense, to speak of "the university" as a singular, integrated institution. Imperative in the sense that "the university" retains both rhetorical and functional significance as a conceptual category. The university as an institution continues to occupy a distinctive place in society as the site of meaningful and deep research-based investigation into the world. The university offers important and powerful institutional complexity, richness, and autonomy for itself and for its researchers, teachers, and students. The university remains an institution whose core purpose is the production and distribution of knowledge,[40] even while its economic scale often links those purposes to many others—economic development, employment, athletics and entertainment, and more. The thousand-year history of the university suggests a continuity of tradition and practice in all of these senses, even if the antecedents of the modern research university are intellectually and culturally diverse.

That description of the university as a singular institution with a continuous history is likewise inappropriate. The institutional character of the university evolved over many centuries. It was not given. The evolution continues. The *research* university, as we know it today, evolved relatively recently. The idea that an institution might be designed to produce *new* knowledge, in addition to documenting and sharing *existing* knowledge, had to wait for the moment that enabled the development of modern science as a research-based enterprise. In its pre-Enlightenment forms, the university denoted an institutional arrangement for re-producing knowledge, embodied in the body of the university's faculty and in its graduates. Post-Enlightenment, with the emergence of German research-based universities in the early 1800s and the adoption of German forms in the US toward the latter part of the 19th century, the modern research university came into focus.

In this brief re-telling of that continuous yet contingent history, the Chapter focuses on three themes. One is the centuries-old conceptual framing of the knowledge now embedded in the university environment: basic or pure knowledge as distinct from applied or useful knowledge. Two is the changed institutional implementation of that dualistic framing in modern universities. Three is the contemporary governance modality that now characterizes and advances that dualism.

B. Knowledge through the centuries

[39] *See* Madison, et al., *supra* note 5.
[40] *See*, *e.g.*, JONATHAN R. COLE, TOWARD A MORE PERFECT UNIVERSITY (2016).

Universities produce knowledge. But what sorts of knowledge, and why? From their beginnings in Europe in the 12th and 13th centuries, until the epistemological revolution wrought by the Enlightenment and practiced by members of the so-called "Republic of Letters" through the mid-18th century,[41] universities and university researchers emphasized the study and sharing of pure, basic knowledge—the concept being traceable to its ancient Greek roots. University knowledge broad forward ancient interests in *epistêmê*, referring to timeless, unchanging, truths about both physical and spiritual worlds, rather than *technê*, referring to craft and practice.

Enlightenment epistemology redefined that duality, characterizing basic knowledge as grounded in empiricism rather than reflection.[42] German universities were the first to implement that shift institutionally, organizing and re-organizing themselves around scientific research grounded in observations of the world and around open sharing of research results, principles that are associated with the writing of Francis Bacon. The University of Berlin, established in the early 19th century on the initiative of Wilhelm von Humboldt, is most famously associated with establishing the foundations of the modern university as a home for research, and for the systematic investigation and production of fundamental knowledge. *Wissenschaft* is the well-known German term for this mode of research and institutional design, aligning the ancient interest in basic truths with modern understanding that truth is the product of investigation, not reflection. As that pairing expanded beyond Germany, Cardinal Newman captured it in his classic *The Idea of a University* in 1852. He described a proper university as a place of teaching universal knowledge, a community of thinkers, seeking truth and understanding judgment via both teaching and research, that is, via broad and inclusive educational practices rather than via vocational pursuits or spiritual improvement.[43]

German *Wissenschaft* was distinguished both conceptually and institutionally from technical training, or applied science. German universities were unlike the German polytechnical institutes, the *Technische Hochschulen*. When researchers and higher education leaders in the US looked to Europe for research practices to borrow and build on in advancing science in US, they looked to the former and not to the latter. In the later the 19th century, US recipients of German research degrees took up academic teaching positions on their return to North America; German-style university research programs were models for new

[41] JOEL MOKYR, A CULTURE OF GROWTH: THE ORIGINS OF THE MODERN ECONOMY (2018).
[42] The historical review in this section is based on Stokes, *supra* note 6.
[43] JOHN HENRY NEWMAN, THE IDEA OF A UNIVERSITY DEFINED AND ILLUSTRATED (1852); The centrality and durability of the vision of the university as a home for basic knowledge are illustrated by the fact that both themes are integral to Jaroslav Pelikan's updating of Newman 140 years later. *See* JAROSLAV PELIKAN, THE IDEA OF THE UNIVERSITY: A REEXAMINATION (1992) (updating and examining JOHN HENRY NEWMAN, THE IDEA OF A UNIVERSITY DEFINED AND ILLUSTRATED (1852)); Pelikan, *supra* note 4.

US research universities (particularly John Hopkins University, the University of Chicago, and Stanford University), for new PhD programs at existing US colleges, such as Harvard and Yale, and for the expansive visions adopted by certain universities that benefitted from Morrill Act "land grant" status (particularly the University of California, at Berkeley, and the University of Wisconsin). Basic knowledge based on research, rather than useful or applied knowledge, was the priority and the domain of authentic modern research universities.

C. Knowledge in the modern university

That conceptual and organizational principle carried over throughout the 20th century, but institutional practice—bureaucratic, professional, and financial—rendered the distinction between basic knowledge and applied knowledge far blurrier on the ground. Professional schools expanded alongside PhD programs. Harvard University created the first modern law school in 1870. The University of Pennsylvania opened the Wharton School, for management training, in 1881. Stanford University launched its School of Engineering in 1925. Yale University founded its School of Music in 1894 and its School of Art in 1869. US universities that were known primarily as engineering institutes (such as the Massachusetts Institute of Technology, founded in 1861) or as homes to agricultural and mechanical research and training (such as many "land grant" universities), expanded into research in social sciences and the humanities. In Germany, a relatively bright institutional line divided basic research and knowledge (universities) from applied research and knowledge (technical institutes). In the US, that line was drawn less distinctly, because it was drawn inside each university. Within that university, basic research and knowledge was to be open and shared by scholars; applied and useful research was to be put to use in preparing the professions and supporting industry.

D. Contemporary knowledge governance in the university

Bringing that history up to the present means highlighting two further critical developments, each of which both relied on and advanced the distinctions just traced back through time.

The first development was production toward the end of World War II of the report titled *Science – The Endless Frontier* by Vannevar Bush, Director of the Office of Scientific Research and Development.[44] The report proposed a massive scaling-up of federal sponsorship of basic research to be conducted in the country's universities and research institutes, de-linked from military purposes, and allowed

---

44 VANNEVAR BUSH, SCIENCE –THE ENDLESS FRONTIER (1945)[hereinafter, BUSH, SCIENCE]. Bush was an engineer, former Dean of MIT's School of Engineering, and president of the Carnegie Institution for Science. He is also famed for his authorship of a visionary essay on communications technology. *See* VANNEVAR BUSH, WE MAY THINK, IN THE ATLANTIC (JULY 1945).

to advance almost exclusively under the supervision of autonomous research experts themselves. Bush's suggestions for institutionalizing this vision in the federal bureaucracy, particularly the idea that the research would conducted in a spirit of free and unfettered inquiry, almost entirely free of government oversight, was less successful than its conceptual framework, which has animated US government sponsorship of university research ever since.

Bush clearly and unambiguously distinguished basic or pure research and general knowledge, the domain of scientific researchers to be supported by federal funding, from applied science and useful knowledge, including its commercial applications, which flowed from basic research in a kind of linear progression and which was the domain of industry.[45] He argued that a well-supported and well-structured domain of basic knowledge production was essential to American health, prosperity, and security, particularly as the Cold War dawned, specifically because (he argued) basic research led to applied knowledge and technology. The US grant-making apparatus that eventually emerged after World War II, distributed across the National Science Foundation, the National Institutes of Health, the Department of Defense, and other agencies, relies heavily on Bush's distinction. Basic knowledge was no longer a good in itself (*Wissenschaft*), but instead an input into a critical pathway toward useful knowledge.

The second development was enactment of the federal Bayh-Dole legislation by the US Congress, in 1980.[46] As a matter of formal law and institutional practice, the Bayh-Dole Act standardized US policy regarding ownership of patentable inventions produced by recipients of federal research sponsorship. Previously, US policy was not standardized; different federal agencies used different approaches, leading to concern—in an era of generalized economic anxiety about American economic competitiveness relative to reindustrializing Japan—that federal support for American scientists was not producing significantly commercially useful results. The policy solution was to permit universities and university researchers to retain patent rights to their inventions, even if the research was conducted with federal support. The Bayh-Dole Act and successor legislation almost single-handedly gave rise to modern university-based technology transfer practice, as universities came to rely on their patent interests to obtain patents, build portfolios of licensed technology, and try to benefit from equity and income from spin-off companies. Technology transfer offices in research universities operate today as boundary worlds, translating ideologically and practically between the applied research practices of the university and the product development and financing practices of the commercial market.[47]

---

[45] BUSH, SCIENCE, *supra* note 44, at 75.
[46] Patent and Trademark Law Amendments Act, 96 P.L. 517, 94 Stat. 3015 (Dec. 12, 1980).
[47] Brett M. Frischmann, Michael J. Madison, & Katherine J. Strandburg, *Governing Knowledge Commons*, *in* GOVERNING KNOWLEDGE COMMONS 1, 26 (Frischmann, Madison, & Strandburg eds., 2014) (citing Patrick L. Jones & Katherine J. Strandburg, *Technology Transfer and an Information View of Universities: A Conceptual Framework for*

The Bayh-Dole Act and modern technology transfer reified the long-standing divide described earlier between basic knowledge and applied or useful knowledge, assumed as a premise that the importance of basic knowledge lies in its capacity for producing useful knowledge, and baked both arguments into the structure of modern IP law. Basic research and knowledge were to be financed largely by the government as forms of what economists call public goods; applied and useful research were to be distributed through society as patented private goods, in the marketplace. Virtually all US research universities now pair their technology transfer organizations with university-wide governance policies that distinguish between copyright law and patent law. Details vary from university to university, but in broad terms, scholarship is the domain of copyright law, and the normative structure of copyright law encourages openness and broad distribution of scholarly knowledge. Invention is the domain of patent law and proprietary markets.

In practice, that means that scholarly works, such as journal articles and monographs, are governed by copyright but are typically remitted by university governance frameworks to the control of their researcher authors. Governance is largely normative. Researcher scholars are expected to disseminate their research results as widely as possible, carrying on the long-standing "Republic of Science" tradition of open sharing. In practice, many if not all scholarly publishers are for-profit enterprises that demand that researchers turn over their copyrights as a condition of publication, giving rise to substantial conflict about proprietary systems and Open Access alternatives. But the normative premise is clear: basic knowledge should be open knowledge.

Useful knowledge, known in modern culture as technology or inventions, may be located within research results or may derive from them. In either case, useful knowledge is governed by patent. Within the university, governance frameworks for inventions are largely dictated by the requirements of the Bayh-Dole Act, which means that federally-supported inventions must be disclosed by researchers to the university so that the university may choose whether or not to claim ownership of them via the patenting process. Governance is highly bureaucratic. Bayh-Dole, technology transfer, and patent law combine to pull researchers toward applied research.[48] Researcher inventors are expected to (and the more entrepreneurial researchers often volunteer to) limit dissemination of their work product other than through the closed mechanisms of the market. The normative premise is uncontroversial: useful knowledge may be closed knowledge.

The foregoing summary is necessarily crude, and readers will see that it is more effective as a description of intellectual history than as an accurate account

---

Academic Freedom, IP, Technology Transfer and the University Mission (unpublished manuscript) (Feb. 22, 2010)).

[48] *See* Brett M. Frischmann, *The Pull of Patents*, 77 FORDHAM L. REV. 2143 (2009).

of scientific practice. Scientific research specifically and empirical investigation of the world generally often do not assert or build on clear distinctions between basic and applied knowledge, no matter what the Greeks believed, or on a linear relationship between the two.[49] But ideas matter, and these ideas—contrasts between basic knowledge in open institutions, and applied knowledge in controlled institutions—have important implications. Two of these are explanatory, having to do with institutional diversity and with contemporary critiques of the university. The third brings the discussion back to data, datasets, and data governance.

One implication is that institutional pluralism in modern universities owes a lot to assumptions and practices built on distinctions between basic and applied or useful knowledge, open and proprietary research results, and applications of copyright law versus patent law. Distinctions between "public" and "private" US universities today are often less significant, in terms of their knowledge functions, than their classification as "research-intensive" or other. Similarly, "elite" and "non-elite" designations are often trumped by the scale of a university's research program and research funding. In turn, differences among universities on those dimensions explain a great deal of the practices of specific universities with respect to investment in research infrastructures and the scope and types of field-specific research practices, needs, and goals. Outside the US, the role of this distinction in institutional pluralism is even clearer, because scientific research has continued to rely on funding practices and institutional arrangements that are more closely linked to national research programs and scientific institutes and less directly linked to the university.

In short, the knowledge production and distribution functions of each university are embedded in specific, local institutional cultures and programs. No two universities are precisely alike. But the synthesis of humanities researchers and scientific researchers in specific universities, and how the university provisions them, within the university's umbrella of unitary institutions, is explained largely by the character of the university's commitment to independent knowledge production. As to governance, the university makes possible the creation of disciplinary communities and fields, sometimes referred to as communities of practice.[50] Self-governance of the university at the local level is often one of its most critical defining features. What counts as knowledge and scholarship within a given discipline is always changing, at least modestly, but is fundamentally a question of expert determination within the discipline. As to provisions, functionally, within its subsidiary schools, colleges, departments, and faculties, and across those units from institution to institution, the university

---

49 THOMAS S. KUHN, THE STRUCTURE OF SCIENTIFIC REVOLUTIONS (3d ed. 1996 (1962)).
50 PAUL DUGUID, COMMUNITY OF PRACTICE THEN AND NOW, IN ORGANIZING FOR THE CREATIVE ECONOMY: COMMUNITY, PRACTICE, AND CAPITALISM (Ash Amin & Joanne Roberts eds., 2008); ANDREW ABBOTT, DEPARTMENT AND DISCIPLINE: CHICAGO SOCIOLOGY AT ONE HUNDRED (1999).

supplies various material infrastructures for documenting, distributing, and collecting knowledge products across all domains, primarily theses, papers, books, and the contents of lectures and tutorials. Capital-intensive research programs depend on expensive devices, laboratories, and related research environments, much of which is funded by blends of internal and external support. University libraries, university presses, and university-based journals support scholarship across multiple fields. In total, as convener and coordinator, the university offers useful and efficient organizational and physical means for scholars of similar stripes to gather in a single place and for teachers to gather with students. Gathering is equal part gatekeeping. The distinctiveness of the university depends, often, on how it imposes admission and membership requirements on both researchers and students.

The second implication is that the university is apt to be called to account by critics either for its failure to adhere appropriately to its ancient anchoring in pure knowledge, or for its failure to abide by the basic/applied knowledge binary described earlier, or both. The university has come to be valuable socially, culturally, economically, and even politically in all sorts of complex ways. That value, or values, entail delicate contemporary blends of openness and public benefit, on the one hand, and privatization and enclosure, on the other hand, and it compromises historic blends. The first broad, clear modern effort to call attention to conflicts between the university's culturally public character and demands to privatize it in different respects came from Clark Kerr, Chancellor of the University of California Berkeley and later President of the University of California. He argued that the 20th century US university had been charged with so many different missions—economic development for the community, entertainment and social advancement and jobs for graduates, symbolic capital for alumni (often related to athletics)—that he came to reject the "university" label and semi-seriously suggested substituting the word "multiversity."[51]

More recent critiques charge the university with "academic capitalism,"[52] because of the university's current alignment with the expectations of private sector partners and funders. Technology transfer practices and the demands and expectations of IP law in the modern economy are often at the core of the complaint; the university is accused of having abandoned its historical character as a home for the production of pure (or mostly pure) knowledge and having sold out to market merchants.[53] Related critiques look less to engagement between the university and commercial interests in the marketplace and more to entanglement

[51] CLARK KERR, THE USES OF THE UNIVERSITY (5th ed., 2005 (1963)).
[52] *See* SHEILA SLAUGHTER & GARY RHOADES, ACADEMIC CAPITALISM AND THE NEW ECONOMY: MARKETS, STATE, AND HIGHER EDUCATION (2009).
[53] For a thorough account of how the cultural and practical imperatives associated with IP law have compromised the traditional cultural identity of the university, *see* JACOB H. ROOKSBY, THE BRANDING OF THE AMERICAN MIND: HOW UNIVERSITIES CAPTURE, MANAGE, AND MONETIZE IP AND WHY IT MATTERS 178 (2016).

between the university and the state. Here, the charge is often not that the university's basic knowledge character has been polluted by commerce but that the university has gotten entangled in the blurring of boundaries between the market and the state. Critics charge the university with both reflecting and advancing creeping neoliberalism, as both private and public funders monitor and assess the university and its citizens and subordinate poorly-defined public goals ("knowledge") to the dictates of the market.[54] Once a semi-autonomous institution that was largely culturally exempt from regulatory oversight associated with private activity, the university is turned into a legible and regulable environment,[55] often to promote or protect important individual or collective interests (such as access for students), but at the cost of eroding some of the university's distinctiveness.

The critics, in short, do not reject the complex knowledge-based foundations of the university or the dual knowledge governance frameworks that have been built on those foundations. Instead, both university celebrants and skeptics largely reinforce and amplify what has come before.

E. How data illuminates changes and conflicts about the mission of the university

That judgment raises the stakes of the third and most important implication of the knowledge-based model of the modern research university. It illuminates the challenges associated with data, datasets, and data governance. As noted earlier, virtually all research universities today have technology transfer practices and formal policy instruments that address allocation of IP interests in research products and scholarly works across the institution. Technology transfer practices manage the boundary between presumptively open, shared knowledge practices in the scholarly setting and presumptively controlled, proprietary knowledge production (*i.e.*, patent-based practices) in the industrial or commercial setting.[56] Inside the university, this is a fragile peace but a mostly workable one.

Most of those formal university policies and governance declarations, especially in the US, do not address interests in data or datasets. Where and how might data fit in university knowledge governance?

One way to look at the question is to start with the long-standing university commitments to research in basic knowledge. Those practices are closely aligned

---

[54] *See* HENRY A. GIROUX, NEOLIBERALISM'S WAR ON HIGHER EDUCATION (2014).
[55] *See* JAMES C. SCOTT, SEEING LIKE A STATE: HOW CERTAIN SCHEMES TO IMPROVE THE HUMAN CONDITION HAVE FAILED (1999).
[56] *See* KATHERINE J. STRANDBURG, CURIOSITY-DRIVEN RESEARCH AND UNIVERSITY TECHNOLOGY TRANSFER, IN UNIVERSITY ENTREPRENEURSHIP AND TECHNOLOGY TRANSFER: PROCESS, DESIGN, AND IP 93 (Gary D. Libecap ed., 2005).

with normative and pragmatic policies promoting openness and sharing of research results. So long as data gathering, production, and analysis are fairly characterized as basic research, leading to generalized knowledge and shareable results, then historical norms align relatively easily with modern expectations. Designing a corresponding governance instrument ought to be relatively straightforward.

But there are other considerations. Data and datasets are not necessarily so readily and uniformly characterized as modes of basic research and knowledge. They may not be knowledge inputs or knowledge products; they may be knowledge tools and knowledge flows. Modern universities have existing formal and normative knowledge governance frameworks, but those are largely directed to knowledge products. University technology transfer practices and IP policies appear to define the full scope of the relevant governance landscape. As knowledge products, data and datasets might be squeezed into those policies and practices, or closed analogous policies and practice might be developed. So, to the extent that data and datasets have technological utility or even some commercial value, then in some respects they might be candidates for governance treatment by analogy to governance of patentable inventions—even if datasets are not patentable themselves. Or, independent of their scholarly character, data and datasets might resemble intellectual and scholarly works of the sort that are typically governed by copyright. To the extent that contemporary knowledge governance expresses an "open v. closed" framing, that outcome might be normatively preferable. Alignment between datasets in the university and university copyright policies for scholarly works might advance the interests of those who advocate for greater openness and sharing of university-based knowledge.

But that outcome would come at a potential cost, because it would treat data resources for governance purposes as taking fixed forms that may conflict with how data are generated and used in "flow" or "lifecycle" practices and processes, as described earlier. Attaching copyright interests to things to which they ordinarily do not attach, even in the interests of openness, raises both conceptual and practical concerns with respect to the purposes of IP law and broad questions of social, cultural, and economic benefit.[57]

In sum, one can analyze questions about data governance in the university in a relatively straightforward way by relying on the well-established conceptual framework that institutionalizes practices regarding research products in two overlapping and usually consistent ways: along a spectrum that runs from basic, general knowledge to applied, useful knowledge, and along a spectrum that runs from copyright and openness to patent and closure. Those two spectra largely define the institutional character of the modern research university.

---

57 *See* Niva Elkin-Koren, *What Contracts Can't Do: The Limits of Private Ordering in Facilitating a Creative Commons*, 74 FORDHAM L. REV. 375 (2005).

Using that approach, questions of data governance align poorly with the standard spectra. Data are not obviously or always basic knowledge or applied or useful knowledge. In many respects normative considerations suggest that data should be open and shared, but the typical governance hook for that approach is unsatisfying at best and risky at worst. The Chapter offers the hypothesis that by questioning data governance, we open the door to the possibility that the centuries-old knowledge-based framing of the university has been exhausted, and the university's knowledge governance framework grounded in technology transfer and IP law has reached the limits of its utility. Data and datasets are neither useful things nor scholarly things; in many respects data are not things themselves. But neither are data and datasets completely *not* things. The university needs to begin again, in part taking data as a central part of its knowledge mission and in part re-thinking that mission itself. It needs to address data governance questions from the ground up, in institutional context.[58] What is the emerging data-intensive university for?

IV. Data governance and the emerging university, re-framed

The future of the university is a broad, conceptual topic. The design of data governance is a concrete, institution-specific problem. This Part bridges the two, by specifying a series of domains where they are can be assessed concurrently and progress made toward possible solutions simultaneously, without necessarily anchoring the analysis either in attachment to forms of knowledge as such (a legacy conceptual framework) or in technology transfer and IP institutions (a legacy institutional framework).

The neologism "data-intensive university" is itself a speculative response to the hypothesis that the centuries-old knowledge-based framing of the university has been exhausted, and the university's knowledge governance framework grounded in technology transfer and IP law has reached the limits of its utility.[59] Fleshing out that argument, the Chapter argues that what may unify the emerging university is data itself, practices and programs of collecting, documenting, understanding, using, and caring for evidence, across all fields, domains, and disciplines of inquiry, both existing and new, including arts and culture, science and technology, social sciences and the professions. "Understanding the world and stewarding and improving its condition" would take the place of "universal and then useful knowledge" as the governing paradigm for institutional practice.

This is meant to be a generous and broadly inclusive definition. "Data" are not products or processes only in natural, physical, or biological sciences. Data, as

---

[58] *See* HENRY PLOTKIN, DARWIN MACHINES AND THE NATURE OF KNOWLEDGE (1997).

[59] The neologism is not entirely novel, though other recent uses of the phrase appear to be motivated by interests related to those that animate this Chapter. *See* David M. Berry, *The Data-Intensive University*, STUNLAW (Sept. 14, 2018), http://stunlaw.blogspot.com/2018/09/the-data-intensive-university.html.

evidence, matter in all fields of research. In some, as noted earlier, the rhetoric of "data" may be displaced by processes of embedding practices in different vocabularies and syntax. What counts as evidence in any particular domain may be contested and may be judged in various ways. Moving from "knowledge" as a central narrative of the university to "evidence" as a central narrative is therefore largely a rhetorical adjustment, rather than one motivated by a strong interest in including new programs in universities or excluding existing ones. But the rhetorical advantage would be this, in institutional terms: the gradual demise of the priority long given to distinctions between basic knowledge and applied knowledge and corresponding implementations in knowledge governance practices that embed choice between "open" and "closed." Evidence, meaning data and datasets, would be subsumed within a university governance framework that blends interests in data forms and data flows in institutionally- and contextually-appropriate norms governing data sharing and data exclusivity. On the ground, data governance policies and practices would be anchored in a fundamental commitment to evidence, rather than anchored in the practicalities of a fundamental commitments to one sort of knowledge or another.

Is this possible? Is it desirable? What difference(s) might it make? Both as process and as result, consider the transition via brief reviews of three intersecting perspectives, each of which illustrates both opportunities and barriers to be expected. The first confronts IP itself as a knowledge governance paradigm. The second is a direct exploration of normative claims regarding openness and data. The third examines the purpose of the university as a product of stakeholder interests.

A. The university as an IP institution

As noted earlier, knowledge governance in the university is most often expressed institutionally in terms of the two default regimes of IP law, copyright and patent. Research universities typically have adopted formal policy instruments that address allocation of IP interests in research products and scholarly works across the institution. The point to be elaborated here is how that governance framing tends either to exclude data and datasets, making the dominant IP frameworks either significantly under inclusive with respect to a huge and critically important knowledge resource, or misleading, if those frameworks are deployed in ways that might be stretched to include data.

As to patent law, patentable inventions created by US university researchers are ordinarily expected to be assigned to the university for consideration for patenting and eventually transfer into commercially marketable technologies and related companies. Under the Bayh-Dole Act and related federal regulations, that expectation is reinforced by legal obligations to disclose and assign interests in inventions with respect to work supported by federal research sponsorship.[60] But universities may and sometimes do impose disclosure and assignment obligations on researchers (including students) as a matter of university policy, exceeding what is required by law. Universities have sustained their investments in technology transfer practices in part based on an ideological commitment to the idea of the university as a participant in private markets for technology innovation.[61]

As to copyright law, copyrighted works created by university employees, even scholarly works produced with US federal support, are not required or expected by law to be transferred to the university for possible commercial exploitation. (Computer software, or code, represents an unclear middle ground, because both copyright interests and patent interests may apply.) Under US law, faculty-created copyrighted works might be considered to be university-owned "works made for hire." Still, many universities have adopted policies and practices that formally or informally waive university copyright claims, grant copyrights in faculty-generated works to their authors, or simply announce that faculty authors are copyright owners with respect to their works, and that the university wants little further to do with them. Some universities go farther, disclaiming copyright interests not only in faculty-generated scholarship but also in much faculty-generated teaching materials. The results are far from perfectly clear or consistent,[62] but the overall tenor is to distance the university from copyright in scholarship and to promote open inquiry and open sharing of research results, apart from the commercial market, just as the tenor in patent law is to bring the university into close proximity with it.

The earlier review of legal frameworks applicable to data make it clear that data have no good home in this scheme, nor should they, because they lack the creativity or inventiveness that IP law looks for. Where do data belong? In closed governance frameworks, or open ones? In this "open v. closed" setting, should data be more "like" patented inventions or more "like" copyright works?

Specifically, because the logic of IP is well-established in the university, and because stakeholders invested in university-based IP governance are apt to be stakeholders in data governance as well, there may an understandable tendency in

---

[60] In 2011, the US Supreme Court held that inventions produced by university researchers with federal research funding were not assigned by law to the university. Rather, the university only acquired that title via an express, signed assignment from the inventor. See Bd. of Trs. of the Leland Stanford Junior U. v. Roche Molecular Sys., 563 U.S. 776 (2011). That decision led to a scramble at TTOs nationwide to revise their policies on invention disclosure and assignment by researchers and to ensure that relevant researchers had assignment agreements in place. In 2018, the federal Code of Federal Regulations was revised to require that recipients of federal research funding have assignment obligations in place. 37 C.F.R. § 401.14(f)(2) (2018).

[61] *See* DAVID C. MOWERY, ET AL., IVORY TOWER AND INDUSTRIAL INNOVATION: UNIVERSITY-INDUSTRY TECHNOLOGY TRANSFER BEFORE AND AFTER THE BAYH-DOLE ACT (2004).

[62] See ROOKSBY, *supra* note 53, at 178-205.

the university to approach questions about university-based data governance by analogizing data to IP and by analogizing data governance to IP governance. "Who owns the data?" is a tempting question,[63] as a convenient IP shorthand for asking whether datasets and associated tools amount to "scholarly works" or "scholarly products," whose use and disposition lie in the hands of faculty researchers, or potentially commercializable inventions, whose use and disposition is determined according to university ownership rules and market imperatives that usually dominate technology transfer practice.

There is no correct or even best answer to the question of how to apply IP frameworks in the university by analogy to data. Instead, take note of how IP law is introduced as its overarching governance logic, often presuming that governance begins by asking, "who owns the information?," or "who owns the data?" The question that began as "open v. closed?" from a normative standpoint is revealed as a question that assumes, in both cases, a premise about unilateral ownership and control.

That conclusion is the implication of both the general statement earlier that "knowledge" as an umbrella normative paradigm for the university does not advance contemporary understanding of governance of the university's information resources, and the specific statement earlier that IP-based scholarly openness and technology transfer as institutional practices have been exhausted as a governing logic for university knowledge products, even by analogy, when questions are raised about data. Both statements take as conceptual given that knowledge is a kind of thing, a basic thing or a useful thing, to be transferred into the broader world beyond the library or the laboratory. Data are different. If data are a thing (and commercially, especially, they may be), they may be owned and controlled. Data may not be things at all. The data lifecycle and the idea of data as flow, described earlier, make data resources ill-suited to the ownership-and-control premises of university knowledge governance aligned with IP law. IP law is quite agnostic about data—even skeptical of it. Industry practice now teaches that data may be subjected to contractual regulation, as a pragmatic matter. Data is not ownable nor owned, usually, in research settings, but may be controlled, in part, in commercial settings. If data and datasets are things, they are third sorts of things, neither obviously owned nor unowned, and in many respects they not things at all. As information resources that are fundamental to the practices of the entire research university and that are shared, as flow, in many respects, ownership and control may simply be the wrong question.

B. The normativity of the data-intensive university

The better and more appropriate governance question in the research university, at least, taking data-as-evidence as comprehensive premise and asks

[63] *See* Barbara J. Evans, *Much Ado About Data Ownership*, 25 HARV. J.L. & TECH. 69 (2011).

how to advance that premise through contextually-appropriate rhetorical and institutional practice. One cannot hope to do that in the abstract. One can only hope to do it on the ground, exploring what works and is effective, and what is not. In the research university, there has emerged not only the fact of data-driven and data-intensive research but also a widely shared instinct that much data and dataset practice should be open and shared.

Is it possible to institutionalize that instinct, as matters of practice and policy, while not unduly disrupting inherited practices associated with traditional university knowledge products? The question is what to do about data as an overarching principle and practice, while respecting continuing, critical investments in both scholarship and in technology transfer itself. Organizing the questions in that way implicates all sorts of institutional practices, from recruitment of students to appointment of faculty, researchers, and teaching staff; to professionalized, administrative bureaucracy; to formal policies; to cultural systems of prestige and normative systems of value. These things can be imagined; can they be done?

Formal policies, one of the questions with which this Chapter began, is an area where it is being done, where institutional practice is being reshaped by efforts to express anew the norms associated with research production. Formal policies and practices may exist at the level of the university, so that they transcend particular fields or disciplines; at the sub-university level (within faculties or departments); and also at the trans-university or trans-research institution level (within disciplines, and/or among research sponsors and funders). The norms, however, are largely shared: data accessibility; data sharing; research reproducibility and research transparency; data accuracy and integrity; data security and data privacy; data retention. Those so-called "top level" norms are accompanied by a cluster of related second-order goals, including administrative sustainability for data practices, in terms of enforceability and financial and technical resource allocation; education and cultural anchoring of data-related practices in institutional norms; and compliance with relevant positive law.

Taken together, two generalized, dominant normative visions emerge. "Share data widely and openly" is the first. "Secure data appropriately," against invasion of privacy interests, misattributions of origin and analysis, and misappropriation by non-researchers, is the second. The edict "make the data accessible outside the university" is paired with the edict, "keep the data secure inside the university." Ensure the free flow of research and scholarship, and do not interfere with effective commercialization of inventions. The historical framing of the university in terms of different forms of knowledge is recast, normatively, in terms of the multiple roles played by data.

Three brief examples illustrate these clusters of concepts may be wrangled into institutional practices and policies that have concrete application. Implementing data as the normative premise of the research university may be a product of organizing from below, inductively, rather determining an optimal strategy from above, deductively. The data-intensive university may turn out to be an evolutionary adaptation of the Baconian research paradigm.

Example one is the effort documented by European researchers to advocate for the adoption of the "FAIR" guiding principles for scientific data management and stewardship. "FAIR" is an acronym; in English, it stands for making data and datasets *findable*, *accessible*, *interoperable*, and *reusable*.[64] The point of the principles as that data governance strategies in particular research institutions should express those values relative to digital assets stewarded by those institutions, and they should express those values as institutional policy and as practice. FAIR information practices are slowly gaining adherents and implementation in Europe, in the UK, and in the US as a set of standards and practices adopted voluntarily.

Example two is the Open Science movement, a community of researchers and allies who are building on successes of the Open Access movement for scholarly publishing to advocate for building and implementing institutional infrastructures that promote open, nonproprietary sharing of research results in repositories and elsewhere.[65] The Open Science movement, like the FAIR movement, explicitly adopts a normative stance, and like the FAIR movement it directs its normative claims to institutional actors like universities and funders. A major component of Open Science advocacy is supporting institutions in building and operating repositories for datasets.

Example three is the BD2K initiative at the National Institutes of Health.[66] BD2K stands for Big Data to Knowledge, and it consists of a large-scale, long-term funding program administered by NIH with the express aim of developing a trans-disciplinary infrastructure for managing and making the best uses of Big Data datasets. BD2K grants encourage recipients to adopt and practice FAIR principles, illustrating different ways in which normative arguments may influence the shapes of data governance ecologies. The US experience so far contrasts with the approach taking in the United Kingdom. Researchers who receive funding provided by research councils united under the umbrella United Kingdom Research and Innovation (UKRI) are expected to provide data management and data sharing plans with all grant and fellowship applications.[67] Guidance from research councils

---

[64] *See* Mark D. Wilkinson, et al., *The FAIR Guiding Principles for Scientific Data Management and Stewardship*, 3 Scientific Data, Article number: 160018 (2016), doi:10.1038/sdata.2016.18.
[65] *See Open Science by Design*, *supra* note 2.
[66] *See Office of Strategic Coordination – The Common Fund, Big Data to Knowledge*, NAT'L INST. OF HEALTH, https://commonfund.nih.gov/bd2k.
[67] *Funders' Data Plan Requirements*, DIG. CURATION CTR., http://www.dcc.ac.uk/resources/data-management-plans/funders-requirements.

is aligned with a set of common principles on data sharing provided by UKRI.[68] Many UK universities have voluntarily affirmatively adopted comprehensive data management strategies. In the US, no equivalent governance framework has been adopted across all government-sponsored research funders, though NIH and NSF requirements that researchers submit data management plans with their proposals, described earlier, affect broad domains of university-based research. US universities have not yet moved as a group to adopt institution-wide data management policies. US practice to date consists in part of development of such policies[69] and in part of guidance that individual researchers should comply with funders' requirements.[70]

To be sure, each of these efforts is advancing in bits and pieces rather than consistently at the level of the university itself, and not each of them consistently avoids or successfully addresses potentially problematic attributes of data governance in the university setting. Data governance should be generous, pluralistic, and inclusive with respect to present and future disciplinary scope and research practice; should be open and inclusive with respect to the professional status of researchers and related data science, information science, and informatics professionals; and should avoid replicating problematic bureaucracies and prestige and power hierarchies in universities and in other research institutions. Shifting from "knowledge" in the research university to "data" and "evidence" in the research university should not be an excuse to suppress or bureaucratize curiosity-driven research or innovative research or researchers. Nor should any of the foregoing be read as altering normative and institutional commitments to university students.

C. The university as the sum of its stakeholders

Addressing those questions, and building upward and outward from the examples given and from others to come, entails a host of complex local questions and conversations. What is best for a particular university and its citizens? Some university policies are quite broad, brief, and largely aspirational.[71] Some are more

---

68 *Common Principles on Data Policy*, UK RESEARCH AND INNOVATION, https://www.ukri.org/funding/information-for-award-holders/data-policy/common-principles-on-data-policy/.

69 *Research Data and Materials Policy*, YALE U, Aug. 31, 2017, https://provost.yale.edu/sites/default/files/files/Research%20Data%20Policy%2006-07-2018.pdf

70 *Penn Libraries, Data Planning and Management*, U. PENN., https://guides.library.upenn.edu/data-management (last visited Aug 3, 2018).

71 *Research Data Policy*, U. BATH, Apr. 9, 2014, https://www.bath.ac.uk/corporate-information/research-data-policy/. The University of Bath policy is accompanied by particularly clear and useful practical guidance., *Guide to Research Data Management at the University of Bath*, U. BATH, http://www.bath.ac.uk/research/data/ (last visisted Mar. 28, 2019).

focused, or longer, or more conversational than prescriptive.[72] University-based research is heterogeneous and changing, in field, discipline, and community terms. Universities vary widely in the scope of their research program, funding, research culture, and management sophistication. Some research fields are well-known to be based heavily on collecting and analyzing large datasets (biomedical research, public health, astronomy and astrophysics), practices that impose management and security and privacy problems that may be several orders of magnitude greater than problems associated with the smaller or even hand-curated datasets more typical of much social science research. Some researchers may work with datasets so large and complex that they require commercial storage services and are processed via supercomputers. Others may work with datasets maintained in simple spreadsheets stored on laptop computers. Some may have full-time professional data scientists to manage their data needs and to facilitate collaboration with other researchers. Some may rely on time borrowed from graduate students. Some may partner with centralized or specialized library services. Researchers in some fields may even be unaware that the material they work with, such as the contents of art collections, may be classified as data for governance purposes.

Within a given institution, thoughtful governance requires inventorying the data practices of the university and understanding the university's needs and goals relative to data and related information resources and the roles that data, data governance, and technical support for data are expected to play. What resources should be treated as data for governance purposes? Whose practices, needs, and goals matter to answering that question? It is common to observe "spillovers" or externalities of data production, storage, or use. Many spillovers are good; some are harmful. A substantial portion of relevant spillovers involve collaboration, scholarship, and social value that crosses boundaries between one university and another, and/or between one or more universities and research institutes, and/or even between one or more universities and private enterprises. The rhetoric and practice of openness, regarding data and other information resources, is not limited to scholarship and datasets. In technology transfer practice, the phrase "Open Innovation" captures knowledge sharing practices across company research programs and between industry and the university.[73] At the same time, but in distinct contexts, data sharing can be deeply problematic, even harmful, particularly if data sharing has commercial motivations or impacts, or if it implicates government surveillance and security. What's good for and by

---

[72] *Retention and Maintenance of Research Records and Data: Principles and Frequently Asked Questions ("FAQS")*, HARV., Apr. 12, 2017, https://vpr.harvard.edu/files/ovpr-test/files/research_records_and_data_retention_and_maintenance_guidance_rev_2017.pdf.

[73] HENRY CHESBROUGH, OPEN INNOVATION: THE NEW IMPERATIVE FOR CREATING AND PROFITING FROM TECHNOLOGY (2003).

university researchers may not be good for or by Facebook or other 21st century information platforms.[74]

Thoughtful governance also requires policies and directives that speak intelligibly to multiple stakeholders and that are capable of implementation in multiple places and via multiple methods. Within the university, answers may be expressed centrally, in a formal policy or in one or more statements to the effect that compliance with relevant mandates is expected, or locally, in the expectation that schools, departments, laboratories, other units, and individual researchers will manage their data appropriately, and/or in some combination of both. It is increasingly common, for example, for a data management plan with respect to a research project to nominate a "data steward" or "data custodian" who is responsible for complying with both relevant university policies and third-party obligations and expectations.[75]

In short, data governance requires contextual understanding and institutional practice. These are partly empirical questions, and they are also partly normative and aspirational questions. They point to ways in which data governance concerns data as a shared resource shared across specific communities and organizations, rather than data as a knowledge resource inside a specific university. Researchers collaborate across institutions and sometimes collaborate across disciplines. Researchers change jobs. In what respects does researcher mobility intersect with data governance requirements and aspirations? There is no single answer to that question. A data management policy may describe the extent to which the university and its researchers should anticipate and plan for relevant technology contingencies, including ownership of and access to both computer hardware and code.[76] One may say that the problem of data governance in a particular university is nested within the problem of data governance in scholarly research generally.[77] One should not assume that governance strategies at one nesting level automatically translate with equal effect to other nesting levels. Nor is the university necessarily prone to claiming undue control over data while third parties are necessarily advocates for broader sharing. The data-intensive

[74] In 2014, Facebook was criticized heavily for enlisting users involuntarily in an "emotional contagion" experiment. Evan Selinger & Woodrow Hartzog, *Facebook's Emotional Contagion Study and the Ethical Problem of Co-opted Identity in Mediated Environments Where Users Lack Control*, 12 RES. ETHICS 35 (2016).
[75] *See, e.g.*, Sara Rosenbaum, *Data Governance and Stewardship: Designing Data Stewardship Entities and Advancing Data Access*, 45 HEALTH SERVICES RES. 1442 (2010), doi:10.1111/j.1475-6773.2010.01140.x.
[76] See Borgman, *supra* note 7, at 224-37.
[77] *See* ELINOR OSTROM AND CHARLOTTE HESS, A FRAMEWORK FOR ANALYZING THE KNOWLEDGE COMMONS, IN UNDERSTANDING KNOWLEDGE AS A COMMONS: FROM THEORY TO PRACTICE (Charlotte Hess & Elinor Ostrom eds., 2006); JOSEPH E. STIGLITZ, KNOWLEDGE AS A GLOBAL PUBLIC GOOD, IN GLOBAL PUBLIC GOODS: INTERNATIONAL COOPERATION IN THE 21ST CENTURY (Katherine J. Strandburg, Brett M. Frischmann, & Michael J. Madison eds., 1999).

university does not automatically mean reducing or eliminating technology transfer strategies. "Who owns the data?" is not an irrelevant prompt for interrogating contextually-appropriate governance strategies, even if it is an inappropriate overarching prompt. As described earlier, technology transfer in the university can be understood as a contextually-appropriate mode of information sharing, and its detailed implementation can be described institutionally as a part of a data governance framework.

Data governance at ground level entails understanding the intersections of governance strategies and interests at multiple levels. In both conceptual and pragmatic senses, data is often managed as commons via strategies of openness, sharing, and polycentricity,[78] but with contextually-appropriate elements of proprietary management and exclusivity.

V. Conclusion

This Chapter has two purposes. One, it extends ongoing critiques and assessments of the modalities of knowledge production, transmission, and stewardship that characterize the modern research university, highlighting conflicts and contradictions between assumptions and practices that point toward openness, sharing, and the public benefit, on the one hand, and assumptions and practices that point toward closure, proprietary practice, and private interest, on the other hand. Two, it re-orients those critiques and assessments via exploration of emerging interest in data governance in the university. Building on globalization of research networks and rapid advances in information technology, data-intensive research increasingly dominates many areas of university research practice, so much so that the research university traditionally grounded in the pursuit of knowledge may now be understood more accurately as the data-intensive university, in which data are described broadly and pluralistically to include modes of evidence production, collection, analysis, and curation throughout the research enterprise.

Given that re-characterization of the university, standard governance framings in terms of open vs. closed frameworks, often aligned with typical expressions of IP law, are no longer descriptively or normatively sufficient. Open sharing of information resources, in practices of Open Science, Open Access, FAIR information practices, and proprietary distribution of research products via technology transfer (including Open Innovation practice) are now better understood both conceptually and practically as parts of the ecologies and practices of data in modern research.

---

[78] Context-specific case studies are documented in GOVERNING MEDICAL KNOWLEDGE COMMONS (Katherine J. Strandburg, Brett M. Frischmann, & Michael J. Madison eds., 2017).

So, at a practical level, while data governance policies and practices are being drafted, reviewed, and implemented, questions and challenges arise at several levels simultaneously. Some of those derive from the mostly "ordinary" set of challenges that apply when overlapping and potentially inconsistent interests and goals need to be integrated into a policy instrument in a complex organization. For data governance, what is relevant data? Who is covered by a policy instrument? What are relevant actors entitled to do, or expected to do, with data? In each case, why and how, and how will any associated burdens be absorbed by the university or otherwise accounted for in terms of money, time, and expertise? Those mostly "ordinary" challenges expose some broader, more open-facing questions about the character of the university's institutional setting and mission and about background expectations, needs, and goals concerning intellectual resources. Policies and practices concerning data are still emerging. The Chapter focuses on concern that data will be "thing-ified" if they characterized by analogy to patentable inventions and copyrightable works and therefore subjected presumptively to thinking and practice derived from IP law. Data governance throughout the university offers opportunities to promote a thoughtful reimagining of the purposes and practices of research institutions and a broader understanding of the roles of institutions in information policy generally.