

Global Patterns in Data Protection: Insights from Social Science and Comparative Law

Document submitted in partial fulfilment of the requirements for the degree of DPhil in Information, Communication and the Social Sciences

Christopher Crum
Oxford Internet Institute, Balliol College

Supervisor: Prof. Viktor Mayer-Schönberger

60,074 words

Abstract

While the global data protection project has enhanced individuals' legal control over their personal information, its virtues should be reevaluated. Comparative analysis of data protection rules reveals a rule monoculture: Over time, data protection rules have become statistically and legally indistinct through a process of convergence. The assumption that monoculture reflects observational consensus that EU-style rules are best is unsound; data protection monoculture is unlikely to have arisen because of cross-jurisdictional observation and learning. Preserving and promoting rule diversity should be the goal going forward as diversity may facilitate adaptation to a changing regulatory environment.

Chapter 1 begins by conceptualizing rule monoculture and problematizing the EU model. Chapter 2 outlines the historical background of data protection, as well as extant literature on comparative data protection, noting a lack of empirical analysis. Chapter 3 uses computational text analysis and case studies to reveal a degree of statistical and semantic similarity across rules consistent with the existence of a rule monoculture centered on an EU-based model. Chapter 3 simultaneously uncovers jurisdictions whose texts appear to "group" around regionally-specific variations on the monoculture model. Chapter 4 critiques existing theories which might explain the empirical reality uncovered in Chapter 3, demonstrating that they lack falsifiability and do not explain grouping. Chapter 5 establishes that the data protection monoculture is unlikely to be the result of learning. Instead, rapid diffusion, textual similarity consonant with extensive copying, inefficient information networks, and knowledge networks predisposed to EU-style rules, suggest diffusion by imitation. Chapter 6 further undermines the plausibility of diffusion-by-learning, employing the natural experiment posed by rules in Gabon, Montana, Tanzania, and Viet Nam to establish that domestic law plausibly drives grouping. Chapter 7 concludes by recommending strategies to restore rule diversity, including rule-agnostic trade agreements and devolving authority for data protection rules.

ABBREVIATIONS

APEC	Asia-Pacific Economic Cooperation
BRICS	Brazil, Russia, India, China, South Africa, Iran, Egypt, Ethiopia, and the UAE
CBPR	Cross-Border Privacy Rules
CCPA	California Consumer Privacy Act
CDP	Comparative Data Protection
CNIL	*Commission Nationale de l'informatique et des libertés*
DPA	Data Protection Authority
DPD	Data Protection Directive [Directive 95/46/EC]
ECOWAS	Economic Community of West African States
FIPs	Fair Information Practice Principles
FRG	Federal Republic of Germany [*Bundesrepublik Deutschland*]
GDPR	General Data Protection Regulation [Regulation (EU) 2016/679]
GDR	German Democratic Republic [*Deutsche Demokratishe Republik*]
GMTV Cases	Gabon, Montana, Tanzania, Viet Nam collectively
GPA	Global Privacy Assembly
GPEN	Global Privacy Enforcement Network
IAPP	International Association of Privacy Professionals
OECD	Organization for Economic Cooperation and Development
OHADA	Organisation pour l'harmonisation en Afrique du droit des affaires
USMCA	United-States-Mexico-Canada Trade Agreement
WorldLII	World Legal Information Institute

ACKNOWLEDGMENTS

I am so grateful for the long list of people without whom this DPhil project would not have been possible. To begin with, I owe an enormous debt to my family: Catherine, Stephen, and Sarah. Their hard work made this project possible. Their support made it feasible. Their love made it worthwhile.

I also wish to thank my supervisor, Viktor Mayer-Schönberger. My academic journey through Oxford was marked by several inflection points. At each point, Viktor believed in me and in my project. I am honored he accepted the role of being my supervisor and honored even more to call him a friend as well as an academic role model. Without his guidance, I would never have gotten to the finish line. He tirelessly offered inspiration, listened to my ideas, and read my work (even when it was bad). Of course, any remaining errors are entirely my own.

The support of the British Government through the Marshall Scholarship program made it possible for me to study at Oxford. I will forever owe a debt of gratitude to the people and organizations who make that program possible, particularly the Association of Marshal Scholars, the former Foreign and Commonwealth Office, Balliol College, Oxford, and the British people. I hope my career embodies the spirit of generosity, friendship, and goodwill that the scholarship embodies. I also owe a huge debt to the Oxford Internet Institute's Shirley Scholarship program, which provided funding to complete my dissertation. Professor Eccles and Professor Taylor thank you for seeing promise in me.

I would also like to thank the many intermediate reviewers of my dissertation. Each of their eyes sharpened the research questions and methods employed herein. Dr. Jonathan Bright and Professor Sandra Wachter were the first external eyes on my project. Their generous feedback made me think hard about how to improve my approach and saved me from many a wrong turn. Professor Victoria Nash and Professor Jeremias Adams-Prassl provided key thoughts and provocations at the crucial confirmation stage. This dissertation benefited tremendously from their input.

My colleagues and friends at the Oxford Internet Institute were an unwavering source of ideas, feedback, direction, and friendship. Hannah, Felix, Michelle, Jakob, Chrissy, Jo, Paul, Liam, Maggie, Nayana, Amanda, and Charlie, thank you. Thank you for listening to my ideas, for helping me navigate Oxford, and for being a source of laughter and camaraderie, especially at a time when the world moved entirely online. Calvin, Manuel, and Felix thank you for making 41 a welcoming place to work. I would be remiss as well not to mention my BCBC crewmates, Ben, Markus, and Raphaël. Here is to early February mornings on the ISIS and just barely outrunning Wolfson in the Gut.

And finally, I owe my many Marshall colleagues a trip to the Lamb and Flag. Jeremy, Mallika, Laura, Andrew, Maddy, Katie, Gabriella, Aneesh, Ararat, Lucy, and Sarah, you guys are an inspiration.

CONTENTS

Global Data Protection at a Crossroads 1
§1.1 Plan of the Dissertation 3
§1.2 Data Protection at 50: More Questions than Answers 5
§1.2.1 The EU Model and Personal Privacy: Problematizing its Features 7
§1.2.2 The EU Model and Personal Privacy: Problematizing Convergence 10
§1.3 Final Introductory Points 13
The Historical Development of Data Protection Law and Comparative Analysis of It 15
§2.1 A Brief History of Data Protection Law 15
§2.2 Comparative Data Protection (CDP): Definition and Development of an Interdisciplinary Field 20
§2.3 The Convergence Story: A Matter of Unexamined Consensus 26
§2.4 Remaining Questions within CDP 27
Textual Homogeneity in Data Protection Laws: Evidence from Statistical Analysis 29
Part I: Text-as-Data 31
§3.1 Similarity in Legislation and Law: Methods Fit and Background 32
§3.2 Methods 34
§3.2.1 Data Collection 34
§3.2.2 Dataset 35
§3.2.3 Cleaning the Data 36
§3.2.4 Similarity Measures and R Environment 37
§3.3 Contextualizing Similarity in Legislation and Law: Benchmarking 40
§3.4 Results 41
§3.4.1 Robustness Checks: Machine Translation and First-Passed as Proxy 42
§3.4.2 Full Corpus Results 45
§3.4.3 Time-Bracketed Results 49
§3.5 Discussion 56
§3.5.1 The “Queen” Model 57
Part II: Data-as-Text 59
§3.6 Case Study 1: Early Divergence between France and Luxembourg 60
§3.7 Case Study 2: Large Market Outliers 62

§3.8 Case Study 3: Groupings, Islands, and Southern Africa 67
§3.9 Conclusion 70
Comparative Data Protection Theories: Assessing Explanations for Rule Monoculture through Realism, Idealism, and Incrementalism 73
§4.1 The Realist Position 75
§4.2 The Idealist Position 80
§4.3 The "Incrementalist" Position 84
§4.4 The Shortcomings of the Tripartite Approach 86
The Improbability of Policy Learning in the Data Protection Space 90
Part I: Learning and the Diffusion of Data Protection Law 90
§5.1 Conceptualizing Learning: What it is and is Not 91
§5.2 Operationalizing Learning 93
§5.3 Comparing Data Protection Diffusion to Learning 94
Part II: Learning and the Structures of Data Protection Diffusion 99
§5.4 The Absence of Network-Structure Analysis in CDP 100
§5.5 Data Protection Epistemic Communities 103
§5.6 Data Collection and Methods 104
§5.7 Results and Analysis 106
§5.7.1 The Absence of Small Worldness 106
§5.7.2 Siloing and the Centrality of Nodes with EU-Philic Views 107
§5.8 Conclusion 113
Explaining The Unexpected Persistence of Grouping: The Plausible Role of Domestic Law 115
§6.1 A Range of Explanations for the Queen's Tines: From CDP to Comparative Law 117
§6.2 The GMTV Cases 125
§6.2.1 Divergent Revealed Goals Through Text and Legal Meaning 127
§6.2.2 Applicability and Processing Principles 128
§6.2.3 Obligations, Cross-border Rules, and Penalties 133
§6.3 Domestic Law as Path Dependence 135
Rule Diversity Restoration and Research Expansion 140
§7.1 50 Years of Data Protection Revisited 140
§7.2 Preserving Rule Diversity 143
§7.3 Future Research Directions 146
Appendix A – Selected Matrix Values 1

Appendix B – Selected Additional Figures 2
Bibliography 5

Chapter I

Global Data Protection at a Crossroads

Monoculture refers to a population composed entirely of the same type of organism. In the natural world, monocultures are exceedingly vulnerable: Pathogens do not need to adapt in order to spread from organism to organism as each member of the population is biologically indistinguishable. (Heslop-Harrison & Schwarzacher, 2007). Biological diversity, however, increases the possibility that some members of the population survive a pathogen, meaning that diverse populations have a chance to increase their resiliency through natural selection. (Birman & Schneider, 2009). By analogy, legal rules exhibit "rule monoculture" where sets of binding rules in multiple jurisdictions approach legal indistinguishability in text and meaning.

Global data protection law exhibits significant amounts of textual indistinguishability. Take the data protection law of Burkina Faso, drafted in French in 2004, and the Argentine Personal Data Protection Act, drafted in Spanish in 2000, on the definition of sensitive data. *A priori*, one might expect the definition of sensitive data to vary. It can theoretically be drafted any number of ways based on what the cultural context defines as sensitive**.** Instead, similarity is glaring. The 2004 Burkinabe legislation defines sensitive data in Article 20 as data relating to, "the health of [a person], ethnic origins, political, philosophical, or religious opinions, trade union membership or morals."[1] The Argentine Act in Chapter I, Section 2 defines sensitive data as data "revealing racial and ethnic origin, political opinions, religious, philosophical or moral beliefs, labor union membership, and information concerning health conditions or sexual habits or behavior."[2]

Why did this happen? Did Burkina Faso consult or copy Argentina? Given their weak trade links and lack of a shared language, that seems unlikely. Were both countries copying a template developed elsewhere, potentially in Europe? That is more plausible. The 1995 EU Data Protection Directive contains a provision in Section III, Article 8 on "special categories of data," indicating that "Member States shall prohibit the processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, trade-union membership, and the processing of data concerning health or sex life." But if Burkina Faso and Argentina were copying Europe, why did Burkina Faso delete the provision

[1] "Sauf derogation prevue par la loi, il est interdit de collecter ou de traiter sans le consetement exprès de la personne concernée, des données à caractère personnel qui sont relatives à la santé de celle-ci ou qui font apparaître les origins raciales, ethniques, les opinions politiques, philosophiques ou religieuses, l'appartenance syndicale ou les moeurs."

[2] "Datos sensibles: Datos personales que revelan origen racial y étnico, opiniones políticas, convicciones religiosas, filosóficas o morales, afiliación sindical e información referente a la salud o a la vida sexual."

relating to the sensitivity of sexual behavior? Why did both countries, seemingly independently, add moral beliefs to the category of sensitive data? There is a rich story of legal evolution and idiosyncratic political choices here yet to be uncovered. Whatever that story is, the ultimate explanation for the convergence in Argentine and Burkinabe legislation intimately relates to the normativity of data protection—its role in the world and whether it can be characterized as good for personal privacy.

The most compelling interpretations are not promising. If, for example, Burkina Faso and Argentina felt compelled to adopt definitions of sensitive data in line with the European definition and in contravention of a locally-preferred definition, the menace of neo-colonialism springs to life. Convergence in Burkina Faso and Argentina may indicate there is little space for local privacy norm development. Moreover, there is the issue of whether the definitions of "sensitive data" are meaningful for the citizens of Argentina and Burkina Faso. Legal transplants from foreign legal cultures are sometimes ineffective because of incoherence with local legal and political cultures. (Bedner, 2013); (Cohn, 2010). Each data protection law's definition of "sensitive" may thus be doing relatively little to improve personal privacy in either country because it does not cohere with Argentine or Burkinabe information norms. (Nissenbaum, 2010).

One might also worry that the amount of policy learning occurring, the outcome-based comparison of existing models, is negligible. First, it may be negligible because Burkina Faso and Argentina are emulating Europe rather than engaging in active policy learning, (Bennett & Howlett, 1992). The absence of policy learning as an explanation for textual indistinguishability further undermines the idea that convergence is good for privacy in either country because it strengthens the relative explanatory power of political-economic motives for domestic adoption. Such motives are often unrelated to concern for privacy or informational self-determination. Furthermore, the future amount of policy learning possible is diminished by the very fact of convergence. Later-adopting jurisdictions looking to define sensitive data lose the ability to compare the experience of Burkina Faso to Argentina since the definitions are practically the same. As Deakin noted in the European data protection context, an essential prerequisite for evolution towards preferable rules in the long run is "local-level diversity, since without diversity, the stock of knowledge and experience on which the learning process depends is necessarily limited in scope." (Deakin et al., 2017, p. 444). The absence of diversity, because it is a precondition for future policy learning, may hinder adaptation to shifting regulatory contexts in much the same way that the absence of biological diversity precludes evolution through natural selection.

And so, at heart, a rule monoculture in data protection law would be suboptimal for two reasons. First, it would lack the capacity to adapt quickly to new circumstances and a changing world. In

a rule monoculture, adaptation is slower than in conditions of diversity because natural experiments do not provide readily observable information on the best adapted rule formulation. Thus, jurisdictions in a rule monoculture must conduct their own experiments and wait for the results to become apparent. It is usually cheaper and easier just to wholesale copy a rule from abroad without conducting such experiments. Rule monoculture has a consequent self-reinforcing effect: A lack of rule diversity spurs further homogeneity. That precludes evolution. Second, rule monoculture would be suboptimal because it is unlikely that fallible, resource-constrained rule-choosers have through luck or skill happened upon the best possible rule. Policy learning has a better chance of approaching the ideal rule in the long-run.

This dissertation brings to light, for the first time, social science evidence that the global data protection project exhibits a rule monoculture. Rules have statistically and legally converged over time towards an EU model or regional approximations of it. That raises the specter of calcification and an inability to adapt quickly. Regrettably, policy learning was not a mechanism behind the emergence of that rule monoculture. Convergence was simply too fast, too complete, and too reliant on networks of knowledge that lauded the EU model to infer that policy learning took place. Resurrecting rule diversity is the way forward.

§1.1 Plan of the Dissertation

To be sure, the global data protection project has undoubtedly increased the degree to which the average global citizen's personal information remains within their control compared to a counterfactual world without rules surrounding the extraction and processing of such information. But if it is true, as this dissertation argues, that the global convergence in data protection rules has led to a textual monoculture that lacks the rule diversity needed to adapt to new technologies and contexts, the global project has fallen short of its aspirations. And if that rule monoculture is unlikely to have arisen through policy learning—what Bennett and Howlett define as lesson drawing about the effectiveness of instruments employed elsewhere (Bennett & Howlett, 1992, p. 289)—there is little reason to champion it as the inevitable outcome of rational deliberation and thus as inherently good for personal privacy. The goals of this dissertation are therefore twofold: (1) to demonstrate the existence and degree of a data protection rule monoculture and (2) to explain and decipher the mechanisms behind its emergence, particularly those mechanisms that are ignored or sidelined in existing literature.

Those goals trace the arc of the dissertation. Chapter 2 begins by examining some necessary historical background. Why, as a historical matter, does data protection spread across borders? What do scholars mean by convergence in data protection law? In answering these questions, Chapter 2 uncovers a surprising dearth of nuance in the debate over the global data protection convergence as well as the

near complete absence of empirical analysis. One of the primary contributions of this dissertation is to provide the missing empirical evidence with which to evaluate theories of convergence. That process begins with Chapter 3, which demonstrates the existence of the rule monoculture using computational text analysis backed up by case studies which show that computational statistical similarity is borne out in legally-relevant text. Most data protection legislation in the world shows degrees of similarity consonant with large amounts of textual borrowing, i.e., global data protection legislation is relatively textually homogenous, even though GDPR has been only minimally copied as a statistical matter. Regional grouping behavior not previously addressed in the literature is also uncovered.

Chapter 4 examines explanations for the data protection monoculture provided in extant literature. An interdisciplinary field, termed comparative data protection, or "CDP," is constructed by combining political science and law review literature. Three theoretical positions each with their own explanations for the state of the global data protection project are charted. First, the idealist view sees the spread of data protection as the result of negotiation and convergence on the EU-model as "best practice." The realist counters that the spread of data protection is due mostly to economic or legal coercion. Finally, the incrementalist ascribes spread to "policy learning." Chapter 4 posits that the CDP literature falls short in several ways. First, it is bereft of meaningful explanations for textual grouping. Second, it relies on the subjective motivations of actors involved in the process of writing domestic data protection rules. CDP theories are, as a result, difficult to falsify.

These shortcomings are addressed starting in Chapter 5, which argues that while it is impossible to discount any given CDP theory entirely as the theories are not mutually exclusive, the chances that policy learning is occurring are low. This further casts doubt on the normative value of the data protection rule monoculture because it suggests it came about through mechanisms unrelated to deliberation. It supports that argument in two ways. First, it undermines the inference that policy learning is occurring by demonstrating textual similarity results reveal relatively quick convergence plus a relatively high percentage of copied text, and that the adoption of data protection frequently precedes common domestic stimuli—all of which are indicative of emulation rather than learning. Second, it shows that the social and knowledge networks through which learning would have occurred are characterized by significant siloing, the absence of "small-world structure," and high centrality of nodes which express EU-style data protection ideas.

Those results, nevertheless, do not address why regional grouping occurs. Chapter 6 provides one possible answer: Domestic law creates path dependent dynamics by influencing the choice architecture a domestic legislator faces in selecting from existing models of data protection. Where

domestic law factors are shared across jurisdictions within the same region, a model developed in the first adopting jurisdiction in the region begins to have a network effect and becomes the text around which other jurisdictions group. Chapter 6, as a result, addresses the counterargument that textual grouping is evidence of learning. Using a case study approach, Chapter 6 also shows that domestic law retains explanatory power even controlling for traditional factors suggested by the comparative law literature on legal transplants including legal history and political economic factors.

Chapter 7 briefly concludes by returning to the normative questions outlined in this introductory chapter. If the absence of policy learning is one reason why the global data protection convergence is suboptimal and unable to respond to change, restoring the diversity that is a necessary precondition for learning would be salutary from the perspective of personal privacy. Chapter 7 explores two approaches to resurrecting rule diversity. The first is law-based: eliminate persistent causes of homogeneity by enshrining rule-agnostic provisions in trade agreements, such as those found in the United States-Canada-Mexico-Agreement ("USMCA"). The second is structure-based: devolve authority for data protection rules to the constituent parts of a state where possible. Each approach comes with tradeoffs.

§1.2 Data Protection at 50: More Questions than Answers

It is crucial to acknowledge that rule monocultures, unlike biological monocultures, are not necessarily suboptimal. If there is a finite number of possible rule formulations (e.g., how many pence are in a pound), the rule formulation is arbitrary (e.g., what side of the street to drive on), trivial (e.g., is a taco a sandwich), or it is indisputable what the "correct" formulation is (e.g., whether murder should be illegal), then a rule monoculture is perfectly acceptable and even ideal. If the correct formulation is known there is simply no need for further experimentation or rule diversity. (Mayer-Schönberger & Gasser, 2024). Thus, the argument that the EU model is the best possible model is a serious counterargument that must be addressed at the outset, especially since many commentators subscribe to the idea that the EU model of data protection is either the best possible expression of rights to privacy and data protection or the closest existing model. (E.g., (Bygrave & Yeung, 2022; Kuner, 2015; Schwartz, 2019)).

One must admit that the argument is attractive. The EU model has a long track record and a large policy community. A recent scholarly accounting noted 157 independent jurisdictions have adopted some form of data protection since Sweden's pioneering Datalagen rules in 1973. (Greenleaf, 2022).

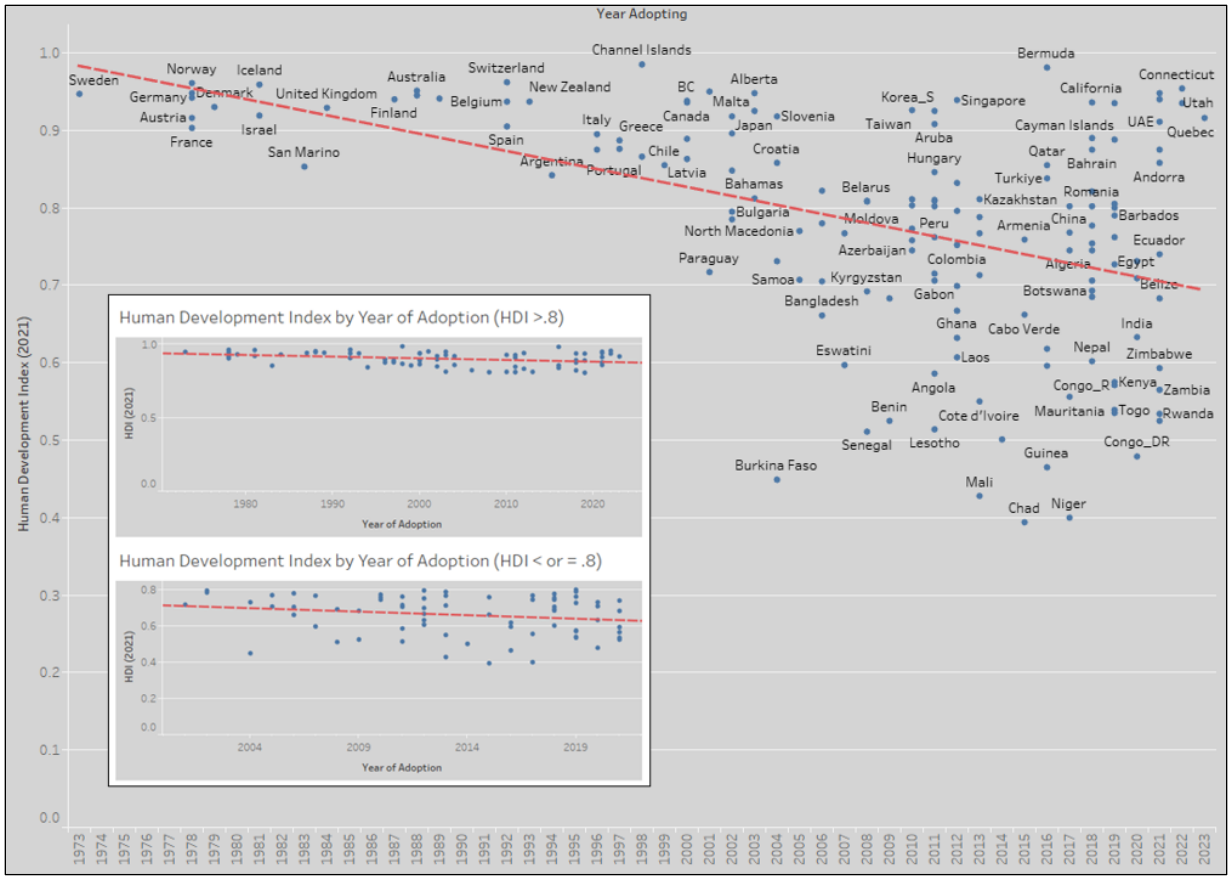


Figure 1.1. Jurisdictions adopting data protection as of March 2023. Human Development Index by Year of Adoption. Red dashed line indicates fitted linear regression (R-squared: .224, p value = .0001). Blue dots indicate jurisdictions. Inset top: HDI by Year of Adoption for jurisdictions with HDI >.8. Red line is fitted linear regression (p value = .001). Blue dots indicate jurisdictions. Inset bottom: HDI by Year of Adoption for jurisdictions with HDI < or = to .8. Red dashed line is fitted linear regression (p value = .103). Blue dots indicate jurisdictions. HDI Data Source = UN Development Program Indices.

By the author's count, 173 of the 193 UN Member States plus Taiwan or about 90 percent of countries in the world not counting independent sub-national jurisdictions have adopted data protection rules as of September 1, 2024.[3] That makes data protection rules more popular than McDonalds (which only exists in 58.5 percent of countries).[4] On a per year basis, between 1973 and 2000, approximately 1.3 jurisdictions per year adopted data protection rules. Between 2000 and 2010, the rate practically tripled to 3.18 per year. Between 2010 and 2020, the rate climbed again to 5.72 per year. The rate climbed yet again to 6 per year between 2020 and March 2023. And since March 2023, 44 additional independent states and subnational jurisdictions formally adopted domestic data protection rules, for a rate of about 27.8 per year. Much, if not most, of the legislation adopted since 1995 is considered to be based on EU

[3] Certain sub-national jurisdictions (e.g., Sint Maarten, Hong Kong) as well as certain rules excluded due to their narrow scope or focus on security rather than data protection as such (e.g., Malawi's Electronic Transactions and Cyber Security Act, 2016) are not in this dissertation's count despite being included in some academic's counts (e.g., (Greenleaf, 2021)).

[4] For context, data protection has yet to approach the popularity of written, codified constitutions (appx. 96 percent according to Hein Online's World Constitution Project).

law, either the 1995 Data Protection Directive ("DPD") or the 2016 General Data Protection Regulation ("GDPR"). Thus, Chander et al., refer to the idea that the EU model is the predominant global model as "the conventional wisdom."(2021, p. 1735). Schwartz argues that jurisdictions have "converged" on EU law. (2019, p. 771). And Greenleaf calls EU law a "global template." (2022, p. 8). How could a model that is the consensus choice for many jurisdictions, a model with a long institutional history and intensely dedicated regulatory community, a model that was updated recently, be anything less than the best possible option?

Despite the consensus, there are reasons to believe the EU model is not the best possible option.[5] There are inherent issues related to its cost, indeterminacy, and dual mandate, i.e., the reality that the EU model aims to facilitate data trade and simultaneously protect privacy. The fact that the EU model has achieved a dominant monoculture exacerbates those issues while also sharply limiting the opportunity for adaptation. There are therefore strong reasons to doubt that an EU-model-based rule monoculture is either better than an alternative rule monoculture in the short run or better than rule diversity over the long run.

§1.2.1 The EU Model and Personal Privacy: Problematizing its Features

Beginning with issues inherent in the model, it is difficult to make the case that the EU model is cost-benefit justified from the perspective of the individual consumer. If one interprets GDPR as a non-tariff trade barrier, by one estimate GDPR costs the average European citizen €260 Euros. (Lee-Makiyama, 2014). While individuals do derive some welfarist value from increased privacy protections—they do not sell their information for nothing—that value is sensitive to context effects (Acquisti et al., 2013), and may not amount to much in real terms. To be sure, increased privacy protections do have positive psychological effects that are hard to quantify in economic terms. (Acquisti et al., 2016, p. 449). Indeed survey data indicate that some consumer fear related to the sanctity of personal information is abated by the presence of data protection law. (CISCO, 2021). But psychological safety may be a double-edged sword if it provides individuals with a false sense of security. And when an individual finds out that in practice they have very little power over their data that creates the conditions for disengagement and mistrust. (Mayer-Schönberger, 2021, p. 3). Actual individual concern for privacy online is down since the pre-GDPR era, (GDMA, 2022), even as threats to privacy have likely remained steady or grown. At least European consumers do appear willing to exercise data protection rights. For example, Google has received 1.39 million requests to delist content since the 2014 ruling in *Google Spain*. (Google

[5] The best option would perfectly align context-relative information norms with information practices (see Nissenbaum (2010)). The best possible option approximates that alignment as closely as possible.

Transparency Report, 2023). Still, the average consumer would have to place a roughly €260 Euro value on the psychological benefit of GDPR plus new personal rights to make EU-style data protection a good value proposition for the average citizen.

The EU model of data protection law is also indeterminate. (Koops, 2014, p. 254); (Bennett & Raab, 2006, p. 125). It is characterized by significant cross-referencing and technicality combined with broad, open-ended terminology that is the result of a political need to compromise. In Europe, that open-endedness allows regulatory agencies to craft the implementation of rules to fit new technological circumstances as they emerge. (Bygrave & Yeung, 2022, pp. 148–152). In practice, however, the open-endedness creates at least two problems: legal endogeneity and prospective uncertainty surrounding new technologies.

Legal endogeneity exists when legal meaning is shaped by regulated entities. (Edelman, 2016, pp. 21–42). It "occurs when compliance professionals on the ground have significant power to define what the law means in practice." (Waldman, 2019, p. 4). The complexity of EU data protection law, its cross references, technicality, and intentional vagueness, supports a cottage industry of lawyers, consultants, engineers, and data protection officers that has grown immensely as data protection rules have spread. The International Association of Privacy Professionals ("IAPP") has more than 50,000 members, (The Privacy Advisor, 2019), and a cursory search of LinkedIn for the title "Chief Privacy Officer" returns more than 78,000 profiles. These compliance professionals, especially CPOs, have significant power to define what the law means by shaping corporate data protection practices and data use policies. (Bamberger & Mulligan, 2015). This is especially true since many Data Protection Authorities have a non-adversarial relationship with regulated entities, often seeking compliance through discussion and awareness raising (Bennett & Raab, 2006, pp. 212–213) as well as "co-regulation," (Hirsch, 2013). When regulated entities play such a large role in shaping what data protection rules mean, in practice the rules take on a managerial, risk-reduction lens. Compliance is "in name only" and often merely "symbolic." (Waldman, 2019, p. 47).

Indeterminacy also creates problems for investment in new technologies. To take the example of self-learning systems, it may be impossible to "make a comprehensive *ex ante* determination of the scope of processing operations conducted by a self-learning" system, rendering data protection hard to apply to many forms of AI, creating significant legal uncertainty for technologists and eroding respect for legal compliance. (Paal, 2022, pp. 307–308). The ongoing standoff between the Italian DPA and OpenAI is symptomatic of this issue.

Furthermore, the EU model may be self-defeating in that it may facilitate the economic conditions which drive the harms it seeks to address. This is in part because the EU model has a "dual mandate;" while it advances fundamental rights, it also seeks to facilitate data trade by reducing cross-border friction and providing legal certainty. (Lynskey, 2015, p. 8). It is counterintuitive but perhaps ultimately unsurprising that data-intensive business models seem to do well under data protection rules. Markets understand this. When jurisdictions pass major data protection legislation, investor confidence in the continued viability of data-intensive business models appears unchanged. The technology-dominated NASDAQ Composite Index, for instance, appears to be untroubled by data protection laws like the GDPR and CCPA (Fig. 1.2). In some cases, it even goes up following passage. The same can be said for the market capitalization of major data economy companies like Alphabet and Amazon (Fig. 1.3). Once GDPR entered into force, the stock prices of Alphabet and Amazon went up (Fig. 1.3).

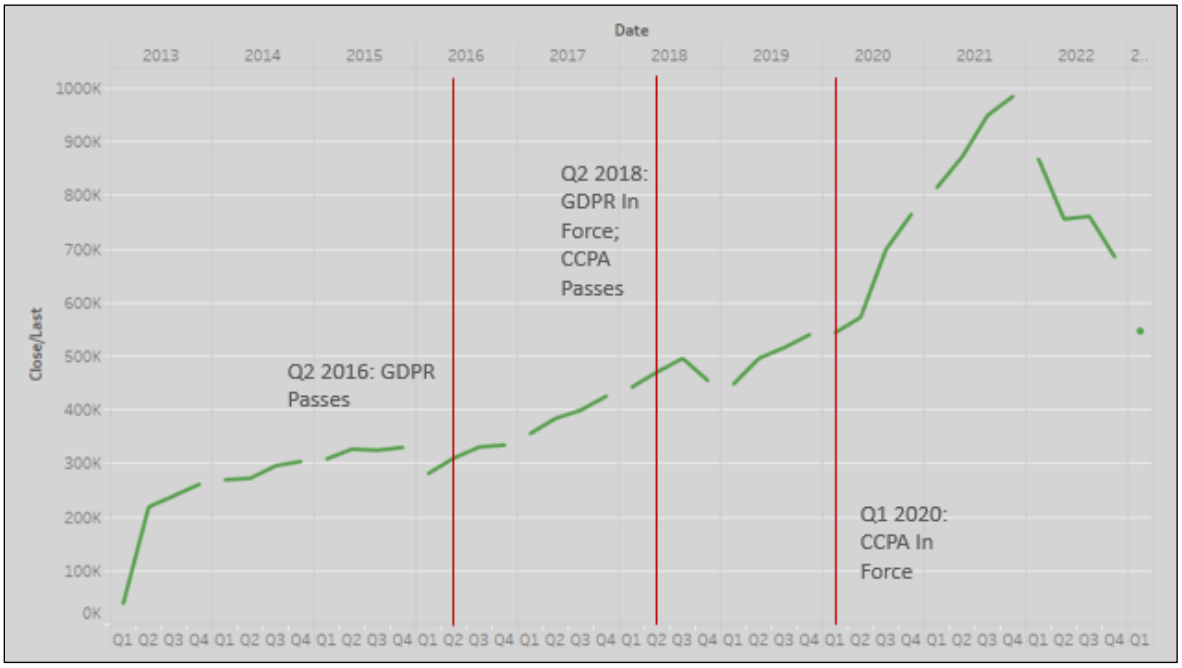


Figure 1.2. NASDAQ Composite Index closing value in USD by date, year and quarter indicated. Red lines indicate passage or entry into force of major data protection legislation. GDPR = General Data Protection Regulation. CCPA = California Consumer Privacy Act. Data source = NASDAQ Market Activity.

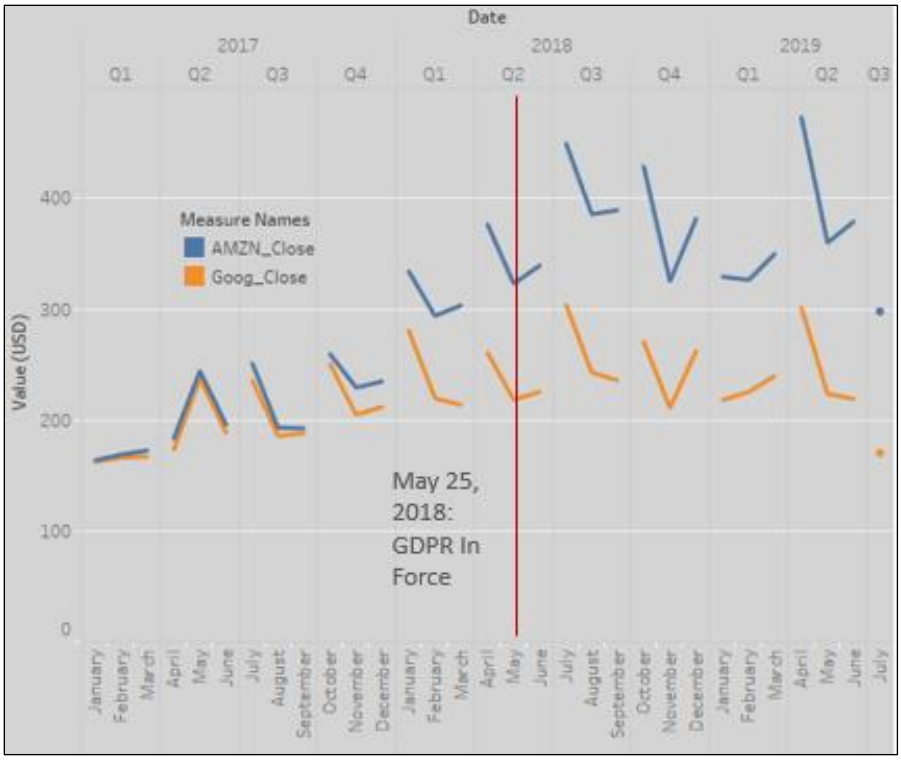


Figure 1.3. Closing stock price on the NASDAQ Stock Exchange of Amazon and Google (now Alphabet) by date, year, quarter, and month indicated. Red line indicates entry into force of the GDPR. Orange line indicates Google closing price in USD. Blue line indicates Amazon closing price in USD. Data source = NASDAQ Market Activity.

At the very least, the growth of data protection has indisputably corresponded with the growth of private economic power. In 1975, around when the first data protection scheme came into force, the world's most profitable company had a market capitalization of roughly $3.33 billion, about one percent of the GDP of the United Kingdom at the time. The market capitalization of Alphabet is $2.02 trillion as of September 2024. Apple Inc.'s market capitalization? $3.49 trillion. Together those two private corporations' market values exceed the GDP of the United Kingdom by about $500 billion. That mammoth economic power translates to power to strategically resist data protection rules, even carefully crafted ones. As Elizabeth Denham, former head of the U.K. Information Commissioner's Office noted to the BBC, "'[one thing] I've learned in this job is that big tech not only has lobbying power but it also has really deep pockets … [their] deep pockets means that there is an inequity of arms." (Rajan, 2021).

§1.2.2 The EU Model and Personal Privacy: Problematizing Convergence

In theory those concerns are mitigated by the fact that global adoption of the EU model gradually eliminates data havens with weak privacy protections and prevents the widespread adoption of a weaker model due to the network effect of its dominant position. (Frankenreiter, 2021; Kuner et al., 2020). But even presuming that is true, global convergence would present issues. To the extent it creates

a rule monoculture, it eliminates productive local variety in privacy rules, variety that forms the basis of the rule diversity necessary to support policy learning and adaptation. Eliminating local variety also crowds out local privacy norms and cultures in a manner that has been characterized as neo-colonialist. (Arora, 2019).

The absence of adaptability would not be an issue in a world that is static, but the regulatory ground is constantly shifting under data protection regulators' collective feet. Technology always races ahead of law.[6] But the geopolitical world is changing too. Power poles are shifting, pandemics are accelerating, and decoupling from international norms and institutions is rampant. (Stephan, 2023, pp. 1–18). As Paul Stephan memorably puts it, "the twenty-first century" is "the place where the hopes of the 1990s went to die." (Stephan, 2023, p. 4). And while proposals to make the digital world personal information-less (e.g., Solid Pods) or radically decentralized (e.g., Web 3.0) might address the conglomeration of data power, they would render data protection as presently designed arguably obsolete. Data protection assumes the digital world is personal data rich, not personal data poor. Bluntly, deep confidence in a data protection model developed in the late 1970s may be misplaced.

Convergence also eliminates vibrant local legal-regulatory privacy cultures and because it occurs around a model created without the input of developing nations, and may have occurred because of economic pressure, could be considered neo-colonialist. Local philosophies of privacy and local privacy norms have much to offer the global data protection project. (Ess, 2006; Reviglio & Alunge, 2020). Unfortunately, as Arora writes, the globalized approach to privacy and data protection law "puts nations in the Global South at a disadvantage as it leaves little scope to institute indigenous privacy norms." (Arora, 2019, p. 369). Threats to privacy which arise within the global periphery and inventive responses to them are sidestepped in favor of wholesale implementation of a model developed in Europe. There is a crowding-out effect for local norms. Local autonomy is imperiled. Even modest convergence in globalized standards "can result in the elimination of the alternatives over which free choice may be exercised effectively" in the future. (Singh Grewal, 2008, p. 292). One could also plausibly worry that data protection norms imposed from outside are less durable. (Berkowitz et al., 2003).

Nor does spread itself necessarily mean personal privacy is increased. Jurisdictions adopt data protection rules for a wide range of reasons including expediency, economic pressure, and domestic

[6] Early data protection scholars understood this and argued for sunset clauses in data protection rules as a consequence. (Simitis, 1987, p. 742).

political interest group preferences.[7] Data is valuable and it behooves a jurisdiction to protect citizens' domestic information from export (so it can be exploited locally) while harmonizing rules internally to give compliant domestic companies a larger data pool to work with. What is more, many of the countries to which data protection has spread do not have a legitimate claim to a human rights or pro-personal privacy mandate. They are frequently illiberal and often undemocratic (Fig. 1.4).

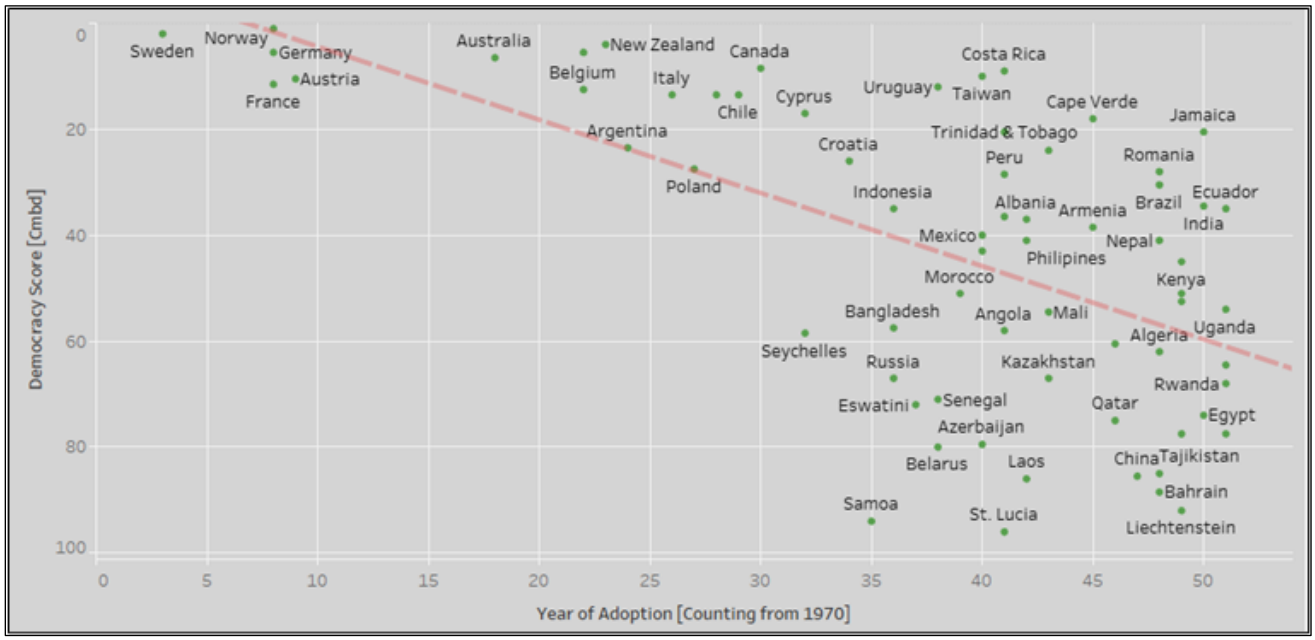


Figure 1.4. Countries adopting data protection as of February 2022, democracy score by year of adoption since 1970. Random subset of 100 countries. Figure does not include non-adopting countries within the sample. Democracy score = average rank within the dataset of country's score on Economist Intelligence Unit's 2020 Democracy Index and University of Würzburg's 2020 Democracy Matrix. E.g., Afghanistan scored a .358 on the Würzburg Democracy Matrix, 61st best in the set, and a 2.85 on the Economist's Democracy Index, 72nd best. Its combined "democracy score" is 66.5. Red dashed line is fitted linear regression (no statistical significance; $p > .05$).

Even assuming good faith adoption, many recently-adopting countries lack enforcement resources. The efficacy of a data protection rule is proportional to its enforcement (Erdos, 2016), and enforcement is immensely expensive. For illustration, the 2022 budget of the European Data Protection Board ("EDPB"), a not-insignificant €6,812,000 Euros, was considered insufficient to carry out all the statutory tasks assigned to the EDPB. (Andrea & Wiewiórowski, 2022). As a result, global spread of data protection rules may instead contribute to the illusion of privacy, something that has been a concern since data protection's early days. (Flaherty, 1989, p. 371). Due to underenforcement on account of resource constraints, privacy *law* rather privacy *protection* may be spreading.

[7] It is also possible that rules spread in a de facto manner as transnational organizations comply with EU rules to save costs and out-compete other market participants. This would unilaterally extend EU rules without any direct coercion. Bradford refers to such a phenomenon as the "Brussels Effect." (Bradford, 2015).

To the extent these concerns are plausible, contemporary data protection rules satisfy *none of the conditions* in which a rule monoculture might be theoretically justified. There are a multitude of ways to write them. They are not arbitrary but instead involve deep socio-political value judgements. They are non-trivial but instead concern billions of dollars in economic value. Most importantly, it should be unclear whether the EU model is the best option. Critical evaluation is therefore vital. If the global convergence on the EU model is suboptimal, why, how, and where has it occurred? That question requires looking back over the past 50 years.

§1.3 Final Introductory Points

But, before launching into history, several further introductory points related to definitions, ethical considerations, data collection, and reproducibility must be addressed.

Law can be an ambiguous term. As used in this dissertation, law refers to rules passed by a legislative body, according to a pre-specified legislative process, that carry with them the prospect of state-backed enforcement and are therefore "binding" on individuals. Statute and law are used roughly synonymously. Judge-made rules of data protection, as well as voluntary codes, are beyond the scope of the project. Nevertheless, the project also deals with "regulation." To ease confusion, this dissertation treats law and regulation as mutually exclusive. When the dissertation refers to data protection "law," it refers to rules passed by a legislature and when it refers to data protection "regulation," it refers to rules passed by any governmental body, other than a legislature, that is empowered to make binding rules. When the dissertation refers to "rules," it is referring to an umbrella category composed of laws and regulations. When it refers to a data protection "regime," it is referring to the cumulative laws, regulations, norms, and institutions in place in a jurisdiction. When the dissertation refers to the "data protection project," it refers to the global cumulative effects of all existing data protection rules.

Data protection herein refers to a subset of rules that regulate the collection and use, *viz.* "processing," of specified data. (Bygrave, 2014, p. 1). Data that does not relate to or reasonably identify a natural person usually falls outside the scope of data protection, (Schwartz & Solove, 2011), and thus outside the scope of the dissertation. While data protection can take the form of soft law, including voluntary codes of conduct, it is usually statutory. (Lynskey, 2015, p. 4). Functionally, data protection statutes have several elements. First, they typically create procedural rules governing the processing of personally identifiable information. Usually, these rules are addressed to organizational entities referred to as "data processors." Second, data protection statutes typically afford rights to the "data subject," the person who the information identifies. Usually, the statutes also create a data protection authority

("DPA") to police the interaction between the "data subject" and the "data processor." (Bygrave, 2014, p. 3; Greenleaf, 2015b, p. 16).

Data protection interests and privacy interests are best conceived of as a Venn Diagram, sometimes they overlap, for example, when a data protection rule prohibits the processing of private personal information, and sometimes they are distinct, as when a data protection rule has no purchase in preventing a physical intrusion into a home that a constitutional rule safeguarding a privacy interest would prevent. This is reflected in EU case law where rights to data protection and rights to privacy are sometimes equivalent, but often separate. (Fuster, 2014, p. 254; Kokott & Sobotta, 2013). So, when the term data protection does not entail privacy interests herein, it is specifically noted.

Some final introductory notes: This dissertation does not employ any personally identifiable data. All data used was publicly accessible (often on a government website). Statistical analysis was conducted in the R statistical computing environment. Code for reproducing results and figures can be found in a dedicated GitHub repository: https://github.com/ccrum1/dissertation. Some figures were produced in Tableau Public, which inherently limits their reproducibility given the software's limited options for saving plots but original screen captures of all Tableau Figures are available from the author along with complete datasets as clean .csv files.

Chapter II

The Historical Development of Data Protection Law and Comparative Analysis of It

The process of looking backward in data protection scholarship and legal development necessitates a unique analytical framework at the intersection of legal and philosophical privacy scholarship, political science, international relations, and global political economy. Without such a multidisciplinary lens, key features of the global data protection project, for example, its inherently transnational scope and aspirations might be missed. Such a lens might be termed comparative data protection or CDP. CDP, as a constructed lens, is historical in that it allows for the periodization of data protection evolution into waves of development, comparative in that it treats data protection as an inherently transnational phenomenon due to the nature of data and rule interdependence between jurisdictions, and normative in that it seeks to assess the value of the global data protection project.

In defining CDP, this chapter serves several purposes. First, it situates the empirical work that follows as part of a third-wave of CDP scholarship. Second, it demonstrates why existing third-wave work lacks the empirical and comparative work necessary to support common CDP views, such as the consensus view that data protection rules are "converging" under conditions of rule interdependence.

§2.1 A Brief History of Data Protection Law

CDP scholarship requires a transnational approach primarily because data protection, from the beginning, was transnational in aspiration and scope leading rule interdependence. The origins of data protection legislation are rooted in the advent of centralized government databases in the 1970s in Western Europe and Scandinavia. These centralized databases created legitimate concerns of the rise of an impenetrable, bureaucratic "surveillance state" catalyzed by advances in computing power. (Flaherty, 1989, p. 373); (Bennett, 1992, p. 2); (Mayer-Schönberger, 1997, p. 221). Early data protection rules sought to exert control over the where and when of these centralized databases and to subject them to defined socio-political values such as transparency and respect for private lives. Since early adopting states, all of which were developed liberal democracies, shared many socio-political values, rules were easily transplanted and adapted across borders, spreading data protection across jurisdictions.

The nature of data was also an early driver of internationalization. From the moment legislation was introduced, the "un-territoriality" of data, to use Daskal's terminology, created problems for data protection, namely, how to ensure the efficacy of a rule-based regime that could be so easily circumvented via export to "data havens" that did not share liberal democratic socio-political values. Thus, even at the earliest stages of data protection development, the inherently transnational nature of

data was clear. (Bing, 1999). As Bing notes, it was "obvious that it would be possible to escape the territory of a state introducing a data protection regime simply by relocating the data to a computer abroad." (Bing, 1999, para. 1). Data are simply too mobile, too indivisible, too location independent, and too intermingled, to be plausibly confined to jurisdictional boundaries. (Daskal, 2015, p. 366); (Chander et al., 2021, p. 1738). Data's characteristics, in short, create "increased capacity and incentives" to abuse personal information whether for private gain or state surveillance. (Reidenberg, 2000, p. 1320). There was a consequent need to either spread data protection rules broadly or ensure data's mobility was limited.

Such exigencies were at first addressed through the reassertion of sovereign control over data export at the national level. For example, under the Datalagen, Sweden's pioneering data protection legislation, personal data held within Sweden could only be transferred abroad with the approval of the Data Inspectorate. (Bing, 1999, para. 2). Similarly, the 1978 French Data Protection act (Act n◦78-17) contains a provision in Article 24 subjecting transfers of data abroad to "prior authorization" by the *Commission nationale de l'informatique et des libertés* ("CNIL"), France's data protection agency. Another example is the German Federal Data Protection Act of 1977; though less explicitly transnational than the Datalagen or Act No. 78-17 in France, it prohibited regulated entities in the private sector from transferring data to internal portions of the entity located abroad without a contract or a showing that the transfer was both necessary and benign. (Riccardi, 1983, Note 195). Unfortunately, these export restrictions did nothing to safeguard the personal information of domestic citizens that was originated abroad and it clashed with the post-war process of building a united Europe best exemplified by the European Communities projects. Consequently, even in a pre-Internet world, the need for robust, internationally agreed on privacy rules was clear to most data protection commentators. (Fuster, 2014, pp. 81–86).

So, by 1980, harmonization was the cause célèbre of data protection advocates. (Bennett, 1997b, p. 227). It was seen as a way to simultaneously universalize privacy protections and remove barriers to transborder data flow created by export restrictions. (Kirby, 1980); (Kirby, 1991). If the rules of many jurisdictions were the same or similar, and provided for a set of universally agreed upon principles of information use and processing, then the threat to personal privacy from personal information originated abroad or exported abroad was significantly lessened, (Gellman, 2022, p. 12), all the while removing legal barriers to data trade by decreasing "conflict" between national regimes (Reidenberg, 2000, p. 1336).

In the early 1980s, two separate international organizations took up the torch of hammering out either binding agreements or soft law guidelines: the Council of Europe and the Organization for Economic Co-operation and Development ("OECD"). (Reidenberg, 2000, p. 1328). The Council of Europe adopted Convention 108 in 1981; it is binding on contracting parties. As of March 2023, 55 nations are contracting parties. The dual purpose of the Convention, to facilitate trade while maintaining personal privacy, is clearly stated in the preamble. Contracting parties adopt Convention 108 "recognizing that it is necessary to reconcile the fundamental values of the respect for privacy *and* [emphasis added] the free flow of information between peoples." (Council of Europe, 1981). Likewise, the 1980 OECD guidelines evince a dual purpose. While the guidelines assert a common interest in privacy among nations, their ultimate purpose is to "further advance the free flow of information," and "to avoid the creation of unjustified obstacles" to the development of "economic and social relations" between member countries. (OECD, 1980). Both international documents operationalize their dual-mandate via fair information practice principles ("FIPs"). (Hartzog, 2017, p. 957). In theory, transnational agreement on FIPs harmonized rules, reduced barriers to trade, and protected personal privacy all at the same time. (Chander & Schwartz, 2023, p. 11).

Both documents, Convention 108 and the 1980 OECD guidelines, illustrate the early internationalization process in data protection, where influence on rules flowed up and down the supranational scale and across jurisdictions. Illustratively, the FIPs adopted by Convention 108 and the OECD Guidelines partly originated with the 1973 Health, Education, and Welfare Report in the United States as well as the report of the Younger Committee on Privacy in the U.K. (Gellman, 2022, pp. 5–6). Moreover, the OECD guidelines and Convention 108 were drawn up following a review of then existing national data protection regimes. (Hartzog, 2017, p. 957); (Mayer-Schönberger, 1997, p. 220). Former Justice of the Australian High Court, and chair of the expert group that drafted the OECD guidelines, Michael Kirby, noted that the national regimes displayed broad agreement on "basic rules," basic rules his group then incorporated in the OECD guidelines. (Kirby, 1980, p. 29). In turn, the FIPs principles were broadly emulated at the national level. (Gellman, 2022, p. 14).

The growth of cross-border data trade brought on by networked communications and the Internet accelerated internationalization. (Reidenberg, 2000). In particular, for the European Union and its political predecessors, entities built in part to dismantle trade barriers, data protection in the digital age "Information Society" presented a "rather obvious problem." (Bing, 1999, para. 6). Enter the Data Protection Directive. The 1995 Data Protection Directive ("DPD") was an EU-level political compromise that sought common ground between the various national-level data protection agencies and regimes,

(Bennett, 1997a, p. 106), particularly between France and Germany, in pursuit of Europe-wide privacy protection and the elimination of national-level idiosyncrasies that had created renewed trade barriers. (Reidenberg, 2000, p. 1335). According to Gellman, "the text of [the Directive] fully reflects FIPs principles, albeit with some variation." (Gellman, 2022, p. 14). The 1995 DPD also reflects the export restriction legacy of early laws. In Article 25 it restricts transfer of data outside the EU unless the third country "ensures an adequate level of protection." As a result, Article 25 functionally exported EU rules to other jurisdictions. (Cate, 1997, pp. 126–127).

Today the General Data Protection Regulation ("GDPR") stands in the shoes of the DPD as a continued attempt to create rule uniformity within Europe while providing a legitimating framework for transnational data flow. (Lynskey, 2015, p. 8). GDPR restricts transfer outside the EU to countries subject to an adequacy decision (GDPR Article 45) or appropriate safeguards. "Appropriate safeguards" include binding contractual agreements in the form of standard clauses set out by the European Commission and binding corporate rules for internal transfers within a transnational entity (GDPR Article 46). As Bennet notes, while the GDPR is a revision of the DPD, "it is also a more multi-faceted instrument, embracing and combining policy instruments that have tended to originate in non-European jurisdictions." (Bennett, 2018, p. 242; Rustad & Koenig, 2019). As such, the GDPR might be thought of as a consolidation of global data protection rules in one place—a global consolidation that scholars views as oft-emulated and more influential than approaches in comparatively-sized jurisdictions. (Schwartz, 2013, p. 1967). For example, many contemporary data protection laws have "adequacy-like" export controls including Saudi Arabia, the UAE, Zimbabwe, Zambia, Sri Lanka, and China. (Greenleaf, 2022, p. 9).

Despite the EU's head start, by the arrival of the GDPR in 2016 Europe was not the only game in town. In 2005, the Asia-Pacific Economic Cooperation ("APEC") established the APEC Privacy Framework and in 2010, the Economic Community of West African States ("ECOWAS") adopted a Supplementary Act on Personal Data Protection. The DPD, however, remained the "most influential" model at the supranational level, prior to the arrival of GDPR. (Greenleaf, 2013, p. 15). The APEC Privacy Framework has not spurred much domestic adoption; it may be "unnecessarily" weak and does not bind any members of APEC. (Greenleaf, 2009). The ECOWAS rules have also had an uneven effect on domestic regimes as domestic regimes have been slow to adopt domestic rules given weak enforcement of the ECOWAS mandate. (Orji, 2017).

Since the mid-2010s, internationalization has continued apace but there was been a renewal of interest in trade restrictions and extraterritorial application of rules. (Svantesson, 2013). This trend had and continues to have two legal aspects: (1) rules for data export and cross border transfer, early

versions of which were surveyed above, and (2) extraterritorial scope of domestic rules. Today, there are two primary statutory methods of addressing/restricting transnational data flow, the adequacy approach, exemplified by the GDPR, and the commitment approach, exemplified by the APEC Cross-Border Privacy Rules ("CBPR"). (Sullivan, 2019). The commitment approach does not require any rules at the national level which might constitute export restrictions. Instead, the commitment approach commits entities that transfer data across borders, whether public or private bodies, to policies that align with the OECD Guidelines. (Rotenberg, 2020, p. 105). These policies must be approved by a designated APEC Accountability Agent. (Rotenberg, 2020, pp. 105–106). These accountability agents can be governmental but are most often private certification and auditing firms. (Sullivan, 2019, p. 384). As one might imagine, the CBPR rules are thus much less complex to comply with than the cross-border transfer rules of the GDPR.

Extraterritorial scope remains a significant part of the internationalization of domestic rules. GDPR, for example, is written to apply to processing activities that take place outside the geographic boundaries of the EU. In pertinent part, Article 3 of the GDPR specifies that GDPR rules apply to the processing of EU citizens' personal data "regardless of whether the processing takes place in the Union or not." GDPR rules also apply to processors located outside the EU when the "processing activities" relate to offering goods and services to EU citizens or monitoring EU citizens' actions within the Union (Article 3(2)). Unsurprisingly, EU rules increasingly conflict with third countries' law. (Kuner, 2015).

Interestingly, the DPD made no such sweeping extraterritorial claims. Websites hosted in third countries, for example, were likely not subject to the DPD even when they collected information on EU citizens. (Moerel, 2011). Instead in Article 4(1)(c), the DPD confined its application to processors located outside the EU to situations where the processor "makes use of equipment … situated on the territory" of a Member State, "unless such equipment is used only for purposes of transit through the territory" of the Union. Many contemporary data protection laws follow the GDPR's lead instead of the DPD. (Greenleaf, 2022, p. 8). Illustratively, the India's proposed 2022 Data Protection Law specifies in Section 4(2) that the act applies "to processing of digital personal data outside the territory of India, if such processing is in connection" with "offering goods or services" to individuals within India. The Oman Personal Data Protection Law, which entered effect in 2023, also follows GDPR; it seemingly applies to any personal data of Omanis processed anywhere at any time without limitation (PDPL Article 2).[8]

[8] Original Arabic:

تسري أحكام هذا القانون على البيانات الشخصية التي تتم معالجتها.

In the battle to acquire jurisdiction over data, many countries have also resorted to reinvigorated export restrictions in the form of data localization rules. Such rules frequently require data, or copies of data, to be located on a server within the jurisdiction in which the data subject holds citizenship. (Chander & Schwartz, 2023, p. 3); (Chander & Lê, 2015, pp. 679–680). Australia, for example, prohibits the transfer of Australian health information abroad. (Chander & Lê, 2015, p. 683). China prevents Chinese Banks from transferring banking data outside the country. (Chander & Lê, 2015, p. 687). And Malaysia requires data about Malaysians to be stored on local servers. (Chander & Lê, 2015, p. 699). As one might expect, calls for restrictions on data flow on privacy grounds are often mixed with motivations related to economic advantage and national sovereignty. (Kuner et al., 2015).

Thus, the modern data protection project confronts many of the same issues that it did in the 1970s. How to protect privacy and subject the gathering of information to democratic control all the while avoiding the creation of trade barriers remains the central question. Internationalization is therefore a push and pull process, where jurisdictions fight to ensure their domestic priorities win out either by enshrining their priorities in international agreements or clawing back the power to regulate data domestically. This push and pull process necessitates a research framework that treats the selection of a particular formulation for a domestic rule or the creation of an international regime (e.g., the OECD guidelines) as a political choice dependent on the choices of other entities in the data protection space. Put differently, the internationalization of data protection requires that CDP view national regimes not as isolated, discrete entities but as a series of rule choices that interact with and affect one another. Political scientists refer to this phenomenon as "rule interdependence." In a globalized world, increasing trade and socio-political ties, whether formal or informal, increase the possibility of rule overlap, a situation where the legal rules of multiple states apply and possibly conflict. (Farrell & Newman, 2015, p. 498). Rule overlap gives rise to rule interdependence, defined as the fact that the rule choices of one state, international organization, or entity seriously impact the costs and benefits of other entities' choices. (Farrell & Newman, 2019, pp. 17–18). Rule interdependence is the backdrop that the study of data protection rules must account for and it is precisely because the world is interdependent that data protection rules diffuse from one national system to another.

§2.2 Comparative Data Protection (CDP): Definition and Development of an Interdisciplinary Field

That realization—the idea that the study of data protection must account for rule interdependence—has two pertinent effects. It means that data protection is productively studied in a global and comparative context. It also means that situating this dissertation and laying out the ideas it draws from entails constructing a *sui generis* field of study. Comparative data protection is an apt name

for such a field. It may be defined as the interdisciplinary study of the global data protection project using the methods of comparative law, political science, international relations, and global political economy. It is comparative, assessing rules not in isolation but in contrast to one another in substance and politics. And, furthermore, it is often normative, seeking to assess the effectiveness of individual data protection rules and the global data protection project in achieving a range of goals including but not limited to personal privacy, trade, data flow, innovation, and user empowerment. Using these constructed contours as guidelines, one can trace the emergence of consensus ideas in CDP and the development of CDP scholarship as well as its shortcomings, particularly the lack of empirical work—a lack this dissertation addresses.

Early works that are precursors of comparative data protection include Hondius (1975), Simitis and Dammann (1977), Bing (1978), Stadlen (1979), Flaherty (1979), Burkert (1981), and Flaherty et al. (1984). Since each adopts a comparative approach, analyzing the differences in national legislation in order to build a sense of what data protection was at the time, this group of works could alternatively be considered the first wave of CDP. Stadlen, for example, compared the legislative provisions of the data protection acts of Austria, Canada, Denmark, France, Norway, Sweden, the U.S.A., and West Germany, finding differences in their respective approaches to public and private sector processing, inclusion or non-inclusion of manual information systems, and approaches to enforcement and transborder data flow. (1979, pp. 176–186). Burkert included Luxembourg and excludes Canada and the United States, but finds significant overlap in features across the surveyed jurisdictions. (1981, pp. 172–176). Flaherty et al. surveyed West Germany, France, Sweden, and the U.S. as well as nascent data protection developments in the U.K. (Flaherty et al., 1984).

Many of the early pre-1980s works retain an underdeveloped sense of what data protection's conceptual relationship to privacy is as well as an underdeveloped sense of the nature of privacy itself. Many of these works cite legal privacy scholarship including Westin (1967), Miller (1971), and Fried (1968). Such early works in the legal philosophy of privacy often reduced privacy to information control (or were perceived to). Westin, for instance, defined privacy as the functional preservation of autonomy, space for self-reflection, and protected communications, within society. Thus, Westin linked privacy to the social and political freedom to define and govern oneself. (Westin, 1967, pp. 23–65). Fried (1968) was even more explicit about privacy as self-governance, i.e., personal control over personal information. He argued that privacy (and implicitly data protection) is "that aspect of social order by

which persons control access to information about themselves." (Fried, 1968, p. 493).[9] While undoubtedly crucial to the development of English-language rights to privacy, these early attempts at a conceptualization of privacy make the early works of CDP dependent on a narrow view of data protection's essence and function. As such, the early comparative works are best employed as historical primary sources indicative of what the field used to look like.

In the late 1980s and early 1990s, especially outside the U.S., the study of data protection shed some of those narrow views by adopting an interdisciplinary approach drawing on political science theories and mixed methods including interviews and case studies. As it did so, comparative data protection took on an evaluative rather than strictly descriptive role. These developments are most conspicuous in English-language literature in Flaherty (1989) and Bennett (1992). Using interviews with key figures in the field, including data protection commissioners, Flaherty was able to go beyond the textual comparison of data protection legislation to evaluate the strengths and weaknesses of the data protection regimes in West Germany, Sweden, France, Canada, and the United States. It was important, on Flaherty's account, to inquire into implementation and enforcement as he noted a "statute by itself is an insufficient countervailing force to the ideological and political pressures for efficiency and monitoring of the population that are at work in Western society." (Flaherty, 1989, p. 381). Flaherty's account also recognizes the importance of politics for the continued success of the data protection project. As he wrote, the success of a data protection regime "depends on the extent to which data protection continues to be a significant political issue." (Flaherty, 1989, p. 407). Without continued political will and legislative interest, data protection legislation would quickly grow stale.

It is Bennett (1992), however, that can make a claim to fully incorporating politics and political science. As he explicitly wrote, data protection is "clearly a policy problem amendable to analysis using the theoretical and methodological tools of the political scientist." (Bennett, 1992, p. 2). Using the tools of "comparative policy analysis" to assess American, British, Swedish, and West German ideas and policies in the data protection space, Bennett concluded that "states have converged around the statutory principles of data protection." (Bennett, 1992, p. 6). However, Bennett found that the policy instruments, essentially the implementation methods in a data protection regime, often diverged in his chosen case studies, necessitating the further importation of political science literature on policy instruments. (Bennett, 1992, p. 7). The import of ideas like "convergence," a "policy toolkit," "path dependence," and explicit comparative casing as to political structure, policy problem, and policy

[9] Informational self-determination remains a common justification for data protection rules. (Thouvenin, 2021).

instrument constituted a step change for CDP. As such, Bennett (1992) marks the rough beginning of a long second wave in the development of the field.

Bennett and Raab (2006) should be considered the apotheosis of the long second wave. Their book is the culmination of the groundwork laid by Flaherty and Bennett in the late 1980s and early 1990s through methodological advances. Once more using the tools of political science, Bennett and Raab surveyed several types of data protection instrument, including sectoral legislation, self-regulation, voluntary codes, and command and control data protection authorities, rather than approach comparison between jurisdictions. At base, they make the argument that evaluating the level of personal privacy may be subjective: "[t]he variety of privacy issues, and the plurality of institutional arenas within which they are addressed, means that different actors can observe progress in one place at one time, and regress in another place at the same or another time." (Bennett & Raab, 2006, p. 294). And they consolidate several key points regarding the global data protection project. They were among the first to assert that "legislation has been prompted more by the desire to reap the commercial and administrative benefits that stem from information technology and data usage than to protect privacy." (2006, p. 241).

The long second wave exemplified by Bennett and Raab (2006) is notable in comparison to early works for employing a plurality of privacy conceptualizations. This is no doubt due to advances in the philosophy of privacy, primarily a recognition that control is an illusory basis on which to ground an interest in data protection, (Gavison, 1980, pp. 349–350), (Tavani, 2007, pp. 7–8), and secondarily, a recognition that privacy has social rather than purely individual value (Regan, 1995, pp. 212–243). A third trend in the philosophy of privacy was a decreased focus on finding a singular "essence" of privacy (Solove, 2002), and an increased focus on defining specific activities that "pose privacy problems" (Solove, 2006).

The long second wave was supplanted by more explicit empirical focus as well as further diversity in topic in the 2010s (e.g., legal discourse and networks), marking the beginning of a third wave of CDP scholarship. On the empirical front, scholars stepped back and asked for the first time, what constitutes a data protection regime, what should count, and how many exist in the world. Greenleaf (2011), the first in a series of Greenelaf papers counting global data protection laws, is a prime example. Rather than take a case study approach, Greenleaf aimed to create an overview of the global picture, and, as he wrote, "the picture that emerges is that data privacy laws are spreading globally, and their number and geographical diversity accelerating."(Greenleaf, 2011, p. 1). Greenleaf counted data protection regimes in 76 separate legal jurisdictions. (2011, p. 1). Along with a comprehensive table of

laws and international commitments to data protection, Greenleaf provided a descriptive narrative of diffusion: legislation originated in Western Europe, before expanding to former Eastern Bloc nations, and spreading to large economies in East Asia and South America. (Greenleaf, 2011). Greenleaf's later work further addressed the empirical picture of the global data protection project. In 2013, he counted 101 data protection rules. (Greenleaf, 2013). In a 2015 update, he counted 109 independent jurisdictions with data protection rules, with European jurisdictions in the minority for the first time. (Greenleaf, 2015a). In 2021, Greenleaf counted 145 jurisdictions. (Greenleaf, 2021). In a 2022 update, Greenleaf counted 157 countries, most "influenced substantially by the EU's GDPR." (Greenleaf, 2022, p. 1). While impressive in its commitments to periodic updates, Greenleaf's counting series left unaddressed detailed analysis of spread beyond tallying rules.

Another aspect of such third wave CDP is a renewed focus on the so-called "Trans-Atlantic Divide." There is consensus that at a high level the EU model of data protection flows from Europe-wide rights charters while the U.S. tends to focus on consumer choice, reifying transparency and transactional fairness, rather than rights. (Chander et al., 2021, p. 1747).[10] Schwartz and Peifer concur. They analyze legal discourse in the U.S. and EU and determine that there is a legal-rhetorical transatlantic divide, concluding that the EU system "protects the individual by granting her fundamental [unalienable] rights," while the U.S. system protects "the individual as a privacy consumer" only against outright deception (Schwartz & Peifer, 2017, pp. 119–121). They call for increased collaboration to bridge the gap. Such works build on canonical work by Whitman, who unearthed a long legal history of privacy as state non-interference in the U.S. and privacy as personal dignity in continental Europe, particularly Germany and France. (Whitman, 2004). Third wave CDP thus points to legal culture and discourse as helpful explanatory variables in comparing data protection regimes—perspectives lacking from earlier work.

Third wave CDP also includes network science perspectives on the diffusion of data protection rules. Bennett introduced "elite networking" as a possible explanation for convergence in the 1990s (Bennett, 1992, p. 5), but the expansion in the number of professional data protection networks involving data protection authorities, consultants, and advocates in the 2010s and beyond brought the network perspective new life. Greenleaf (2013) catalogues at least twelve global and regional associations of data protection authorities dedicated to sharing ideas and assisting in enforcement of rules. (2013, pp. 22–25); (Greenleaf, 2015b). Bennet and Raab refer to the links between DPAs as "an

[10] Sometimes the literature refers to the U.S. model as "notice and consent" to indicate that the consumer only needs notice of what uses personal data will be put to and is then offered a take-it-or-leave-it choice to accept the terms or refrain from using the service. This approach has been universally panned in scholarly literature. (Susser, 2019).

example of bottom-up internationalization" and possibly as evidence of "the early stages of transnational regime-formation." (Bennett & Raab, 2020, p. 453). The network perspective is further exemplified by monographs by Bennett (2010) and Newman (2008). Bennett demonstrates that cross-national networks of advocates and experts are most successful at the agenda-setting phase in defeating privacy-decreasing government measures where they can build a one-off coalition of civil society and consumer protection groups but he remains skeptical about the overall ability of privacy advocates to protect privacy in part due to internal network conflicts. (Bennett, 2010, pp. 169–225). Newman argued that the 1995 DPD and subsequent data protection diffusion from Europe to the world was the result of strong, concentrated networks of data protection regulators within Europe with content expertise sufficient to overcome opposition within the political processes of the EU. (Newman, 2008). The network perspective, especially regarding networks of DPAs and private data protection consultants, is one that is readily susceptible to future research (see, e.g., Chapter 5, Part II).

Contemporary CDP incorporates elements of each wave. It is comparative—e.g., (Sudarwanto & Kharisma, 2022). It rejects a narrow conceptualization of privacy as information control in favor of an issue-specific "family resemblance" or culturally-centered approach—e.g., (Yang & Yan, 2021). And, it is interdisciplinary and concerned with questions of global spread. It also incorporates the global politics of data protection. For example, contemporary CDP unearths attempts in the BRICS countries to create a network of interlocking data protection rules to counter the influence of the EU. (Belli & Doneda, 2022).

Still, contemporary work falls short in several ways. It often focuses on single country studies or issues that miss out on comparative insights—e.g., (Christo, 2013), (Resta, 2014), (Cai & Chen, 2022), (Jung & Joo, 2023), and (Kinikoglu, 2023). As Bennet and Raab caution, "no analysis that confines its frame of reference to only one country or jurisdiction as the field of play… for privacy instruments can be adequate." (Bennett & Raab, 2006, p. 234). Moreover, contemporary work can sometimes be overly legalistic or technical. Either work focuses on how new technology can fit within existing rules rather than how new technological affordances and technological practices might change the rules either functionally or substantively, e.g., (Weitzenboeck et al., 2022), or it focuses on the technical exegesis of existing rules—e.g., (Finck & Pallas, 2020). Additionally, cross citation between journals and across methodological divides is rare. This "siloing" limits the ability of CDP to consolidate first principles and points of agreement. Contemporary CDP has also inherited the tendency of third wave CDP to be empirical only in the simplistic sense of counting rules.

§2.3 The Convergence Story: A Matter of Unexamined Consensus

That lack of empirical work is especially troubling given the consensus in the field that data protection laws are converging across borders, that is progressively getting more similar in meaning and functionality over time, an assumption in need to empirical evidence to back it up. A skeleton "convergence" story—about which there appears to be widespread agreement—can be gleaned from the literature. In the 1970s and 1980s, several Western European countries faced similar problems brought on by the advent of computerized databases. They addressed these problems using a new tool: data protection. "The practical methods" adopted had "more commonality than the national legislation" had differences. (Flaherty, 1989, p. 373). This was in part because "pioneers" had strong incentives to "imitate or adapt" the efforts of other countries in what was then a new policy field. (Bennett, 1992, p. 4). In the 1980s, the OECD Guidelines and Convention 108 further "encouraged" and accelerated a "process of emulation among advanced industrial nations." (Flaherty, 1989, p. 11). As such, the 1970s and 1980s evince a "general pattern" demonstrating "that globalizing forces for economic and technological development were progressively overwhelming the distinctive institutional capacities, historical legacies, and cultural traditions of advanced industrial states." (Bennett, 1997a, p. 100). Agreement on "core principles" was widespread. (Bygrave, 2014, p. 205). Some differences at the level of policy instrument—the enforcement mechanism—remained however. (Reidenberg, 2000, p. 1325). The DPD harmonized some of those differences. "Greater conformity" in policy instrument, especially in Europe, resulted (Bennett, 1997a, p. 118), despite some remaining "idiosyncratic national features." (Reidenberg, 2000, p. 1335).

Then, throughout the 2000s and 2010s, by updating and modernizing the DPD through the GDPR, Europe set itself as the "global standard." (Chander et al., 2021, p. 1735). The GDPR "re-converg[ed]" rules on a shared interpretation of the FIPs "following 20 or more years" in which states' interpretations "had somewhat drifted apart." (Bennett & Raab, 2020, p. 452). Gradually "legal inertia" and European "soft power" became hard to ignore and many lagging countries adopted laws "influenced strongly" by European standards, especially GDPR. (Greenleaf, 2015a, p. 7). As of 2022, GDPR was "clearly the *global template,*" (Greenleaf, 2022, p. 8), albeit a global template with some "variations" in implementation (Greenleaf, 2022, p. 3). Whether influenced by GDPR or independently by the FIPS, which were themselves an influence on GDPR, "most" laws in the 2010s converged on similar principles. (Hartzog, 2017, p. 958). Parallel to the legal convergence, a de facto convergence based on the compliance choices of regulated entities may also have occurred. (Bradford, 2015).

Of course, as the literature cautions, it is important not to overstate the degree of convergence and to be clear about what is converging. Convergence is not a synonym for uniformity. (Bennett, 1992, p. 6). Typically, convergence is thought to have occurred in data protection legislative text and to some degree in policy instrument while policy outcome, due to research difficulties, has not been assessed. (Bennett, 1997a, pp. 119, 120). The interpretation of data protection laws has remained somewhat nationally distinct. (Reidenberg, 2000, p. 1370). Likewise, forces for divergence continue to exist. Government styles and norms are not uniform. (Reidenberg, 2000, p. 1370). The "political and institutional dimensions of the EU's rights talk" discursively move the EU further from the U.S. and vice versa. (Schwartz & Peifer, 2017, p. 120). Furthermore, self-regulatory codes of practice as an approach to data protection tend to entrench some sectoral differences across industries. (Bennett, 1997a, p. 115). In short, contemporary "[c]ross-cutting" norms and interests, e.g., trade, security, rights, and law enforcement, make it "difficult to reach [true] global consensus" on precisely uniform data protection principles. (Bygrave, 2014, p. 206).

At the same time, there may be more than one convergence. Contemporaneous with the global convergence on the GDPR as a legislative template, a subnational, non-Brussels-spurred convergence was occurring in the U.S. (Chander et al., 2021, p. 1737) as U.S. state-level data protection law especially in Connecticut and Colorado "largely copied" the California Consumer Privacy Act, which has a number of non-GDPR elements,[11] rather than GDPR. (Chander et al., 2021, p. 1763).[12]

§2.4 Remaining Questions within CDP

Bringing together multidisciplinary literature on data protection, this chapter named and described CDP as a lens through which to view this dissertation. Literature which fits within the frame supplied by CDP studies data protection under conditions of rule interdependence wrought by the nature of data and the data protection project as internationalized from the start. Agreement within CDP primarily centers on the idea of "convergence," the notion that rules are becoming more similar over time. At some level of generality many states, subnational jurisdictions, and private entities have relatively similar legal regimes of data protection and/or tend to agree on a basic set of principles. Even those like Chander et al. (2021), that dispute aspects of the skeleton story agree that convergences are commonplace in data protection law.

[11] Chander et al. note that "the two laws do not share the same underlying principles, leading to great differences in the scope and nature of the rights and duties imposed by each." (Chander et al., 2021, p. 1755). CCPA also does not require processing to be lawful, which is a massive shift from GDPR's position.

[12] Chapter 3 shows that the idea that Colorado and Connecticut copied California may be suspect, but evidence for state-level convergence is still robust.

What remains a subject of disagreement is precisely what convergence means. What is converging? How similar do two jurisdictions' laws need to be before a convergence is declared? What mechanisms are behind convergence? What model will convergence occur around? Some observers argued that convergence may occur around a weaker standard than the EU model, e.g., the APEC Privacy Framework, resulting in "an incoherent and fragmented patchwork" with "periodic" but "unpredictable victories" for personal privacy. (Bennett & Raab, 2006, p. 295).[13] Others argue that convergence around the EU model will continue unabated until the EU model, or a comparatively stringent model, is ubiquitous. (Schwartz, 2019). There is also the question of if there is a convergence, whose principles win out. Some argue for a "Brussels Effect," where the EU's policy preferences win, (Bradford, 2015), some argue for a "D.C. Effect," where the U.S.'s preferences win (Rustad & Koenig, 2019).

Moreover, whether convergence is a good thing is far from clear. In other words, is the diffusion of data protection rules, " the diffusion of privacy *protection*, or the diffusion of privacy *law*?" (Bennett, 2018, p. 244). Before the dissertation addresses such questions, some more basic questions must be addressed. How high is the degree of convergence? Is convergence indicative of a rule monoculture? What causes convergence? Much empirical work remains to be done.

[13] Greenleaf (2014, p. 559) refers to this as the Bennett-Raab thesis—somewhat inaccurately in the author's opinion. Bennett and Raab's main arguments were that the typology of policy instruments they identified remained stable, should be thought of as interlocking, and was underlined by a shared "privacy paradigm" across jurisdictions. (Bennett & Raab, 2020).

Chapter III

Textual Homogeneity in Data Protection Laws: Evidence from Statistical Analysis

A major step in assessing the veracity and normativity of the convergence story so prominent in CDP literature is to determine patterns of textual similarity over time. If convergence entails degrees of textual similarity indicative of a rule monoculture, that may be deleterious for the success of the global data protection project for the reasons outlined in the Introduction. The convergence story from Chapter 2 implies that text will get progressively more similar over time and that the level of similarity is indicative of agreement, text reuse, borrowing, and/or copying. (Bennett, 1992, p. 6). Normatively, if convergence entails degrees of textual similarity indicative of a rule monoculture, that may be deleterious for the success of the global data protection project for the reasons outlined in the Introduction.

CDP's convergence story is broadly accurate empirically. Over time, data protection texts have gotten more like each other. But what CDP scholarship fails to acknowledge is that convergence processes have in fact created a data protection rule monoculture, the most direct evidence of which is significant textual homogeneity in the global data protection project. While text is not the only aspect of legal meaning, and rule monoculture implies that it is legal meaning that is indistinguishable across jurisdictions, textual similarity is the canary in a coal mine. Rule monoculture is simply unlikely to exist where legal texts are statistically dissimilar. Conversely, where statistical textual similarity exists, legal meaning is likely to be similar as well. Meaning was completely divorced from text is not law but rather rule by edict and data protection is assumed to be a matter of law.

Methodologically, this chapter explores rule monoculture visually, through heatmapping, mathematically, through Jaccard and Cosine Distances, both measures of the statistical similarity of strings, and qualitatively, through case studies motivated by the visual and mathematical analysis. Rather than canvas every piece of legislation on the topic of data protection, it employs a proxy for all legislation passed in a country: the first data protection law passed in a respective jurisdiction. The first law passed is a valuable proxy for a data protection regime for four reasons. (1) The first law passed in a jurisdiction is likely to set the tone and terms for later laws as lawyers, compliance personnel, and most importantly draftspeople, learn to see data protection through the technical terms embedded in the first legislation. (2) Since it is likely cheaper in terms of political capital, time, and staff resources to amend a law rather than start from scratch, the legislative texts of later adopted laws are likely to hew relatively closely to the original law. (3) The omnibus approach to data protection many countries pursue means that later

data protection law merely fills in doctrinal or sectoral gaps in coverage or further specifies rules already laid out in broad strokes in the previously-passed omnibus legislation. (4) In most jurisdictions, the only law passed is the first law passed. Importantly, the first-passed legislation provides an opportunity to assess similarity over time that assessing only the rules on the books now does not. And only by assessing historical versions of rules can convergence as a process rather than a condition be evaluated.

Upon assessment, the first-passed legislation in each jurisdiction is getting more similar over time measured in terms of decreasing Jaccard and Cosine Distances, on average since 1973, confirming the main element of the convergence story. Moreover, as expected, this chapter also uncovers continuing outliers in legislative text, including major economies that rely less on access to the EU data market, e.g., China, the United States, and India. What this chapter further uncovers is that grouping and clustering are also prevalent in the text corpus. What that tends to suggest is that there may be multiple nuclei of convergence or "mini" convergences. One can see this trend most clearly in American states (except California) and Caribbean Island nations as well as some Eastern and Southern African jurisdictions. Just as many countries around the world indeed appear to be getting closer to the EU model, including, in some respects, the American states, Caribbean nations, and African nations in the dataset, the more pronounced trend is that such groups are getting closer to one another.[14]

Structurally, this chapter is split into two conceptual halves. First, in Part I, it uses measures of statistical textual similarity to tell a broad-brush-strokes story about patterns of similarity in data protection legislation over time. These patterns are visualized as color-coded heatmaps, a visualization of matrices consisting of rectangular tiling, with each tile shaded by color to represent a value in the corresponding matrix of statistical textual distance. (Wilkinson & Friendly, 2009, p. 179). Here, darker colors indicate higher degrees of similarity (or more specifically, lower degrees of distance). These heat maps inductively facilitate the identification of case study candidates and patterns for further qualitative inspection, as in Alschner et al. (2018). Qualitative inspection is accomplished in Part II. Part II draws out the implications of the mini convergences and demonstrates that the textual statistical similarity uncovered in Part I corresponds to similarity in legally relevant aspects of legislation.

There a number of ways that similarity can be operationalized. (Shaffer & Elkins, 2019, p. 6). Similarity might be defined as equivalence; texts could be similar in the sense that they use the same words in the same order with the same meaning. Similarity could also mean synonymous; two texts

[14] While it is unlikely this is due to in-group legal or economic coercion, it is not certain that learning is taking place either. Copying and pasting appears to be common. The degree of copying likely implies less reasoned appreciation and more expediency. More detailed analysis of that point is a task for later chapters, in particular Chapters 5 and 5.

could be similar in the sense that they have the same meaning but use different words. Texts could also share the same themes and thus be similar in the sense that they share underlying focus, topic, or elements. Computational assessment of textual similarity typically deals with equivalence as similarity, (Shaffer & Elkins, 2019), and that is the approach taken in Part I which treats similarity as equivalence in text rather than equivalence in meaning.[15] Part II deals with similarity as a matter of similarity in comparative legal meaning by analyzing the similarity of legally-relevant text using a case study approach.

Part I: Text-as-Data

Treating legislative text as data implies that the semantic meanings of the texts themselves are not the analytical substrate. Instead, the text itself is the data and the only question is how statistically similar texts are to one another in different periods of adoption. This section studies textual similarity by benchmarking similarity using known instances of legislative influence and comparing those values to pairs of data protection text by their Jaccard and Cosine Distances. The research task is to assess whether legislative text is converging in the sense that legislative text itself is becoming more statistically similar, defined in terms of textual equivalency, and if so, to assess the patterns present. Which laws are most (and least) similar to each other? How do laws in time brackets compare to one another? How do time brackets compare to each other in terms of similarity over time?

But first Section 3.1 surveys the literature on comparative legal and political science approaches to textual similarity. Section 3.2 then addresses methods, discussing data collection, the dataset itself, the data cleaning process, and finally, the mathematics of the selected similarity measures.[16] Section 3.3 introduces the benchmarking approach. Section 3.4 reports results.

[15] This distinction between text and meaning highlights that computational assessment of textual similarity among legal texts comes in two flavors: syntactic and semantic. (Moodley et al., 2019, p. 66). Semantic approaches parse text as blocks of words and aim to get at meaning either through natural language processing techniques employing machine learning, including word embeddings, or through less computationally intensive approaches that consider sentence-level context (n-grams) but do not attempt to evaluate sentiment or embedded meaning. Syntactic methods, on the other hand, treat individual texts as mere "bags of words." Part I employs n-grams as a simple semantic approach.

[16] Code to reproduce results as well as the .txt files can be accessed at https://github.com/ccrum1/dissertation. Some versions of Windows render certain characters in text files as non-Unicode symbols. This will cause R to encounter a fatal error. If this occurs it may be useful to use an online converter to convert text from the original PDF and only then load it into R. There are several free versions of converter (including PDF2Go and Zamzar). An alternative is to use the lapply function (see markdown file on GitHub) to remove non-Latin characters from text encodings. Both strategies were alternatively employed for this project.

§3.1 Similarity in Legislation and Law: Methods Fit and Background

As a preliminary matter, computational textual similarity, as applied to legal or political text, is a strong fit for assessing the convergence in data protection legislative text. As a method, it works best when manual comparison is impractical given the massive corpus of texts to be compared. Here the 210 legislative texts passed means there are 21,945 unique dyads to evaluate. While not impossible, manual comparison would be prohibitively time consuming and likely to miss instances of similarity across all 21,945 pairs of legislative text.

There are also several theoretical reasons why treating legislation as data is a promising approach here. First, data protection legislative text exhibits an expected balance between textual borrowing and textual invention. Documents that have a standard set of provisions written using similar language tend to be strong candidates to assess using textual similarity. (Alschner & Skougarevskiy, 2016, p. 6). Statistical textual similarity can do very little to compare documents that do not share any terms or phrases; the text-as-data approach will not be able to determine anything about the contents of either text. Data protection legislation, like much legislation in any area of law, tends to share common phrases and writing conventions. For example, many data protection rules have a definitions section defining terms like data controller or data processor. Those shared concepts may mean legislators look abroad to assist them in their drafting task, thus transplanting text from elsewhere. (Barak-Erez, 2009, p. 492). At the same time, jurisdictions usually have local idiosyncrasies in legislative style. In part this is because different drafters have personal style preferences (Bowman, 2006, p. 2), and in part because different jurisdictions have different "stylistic" templates. Some jurisdictions employ a preamble to aid in interpretation, stating findings in the form of whereas clauses, while others have formalized offering clauses, or are phrased as mandates generated by a sovereign. Thus, there is reason to believe that some differences will remain between legislation on the same topic in different jurisdictions, creating rich possibilities for interpretation.

Second, legislative text is deliberate, making text itself analytically significant. Drafters typically understand that a change in language indicates a change in legal meaning. As Curran puts it, legal meaning is "wedded" to the language choices that "encase" them. (Curran, 2019, p. 706). Words are invested with meaning and changes to wordings matter. Words create binding rules: The text in a legislative document is "an aggregate representation of the dimensions underlying" the policies and legal meanings entailed therein. (Linder et al., 2020, p. 549). Lawyers and interpretive professionals, often the very people doing the drafting, are trained to make arguments about legal meaning by comparing the language of similar rules. For all these reasons, "[d]rafters make conscious decisions" on

whether to borrow or depart from “established terms and standardized phrases” knowing it will be a difference the legal audience will find salient (Alschner & Skougarevskiy, 2016, p. 8). As a result, the choice to include or not include template language says something about the drafting intention of a particular legislative text in comparison to another.

To date, there is no existing computational analysis of text reuse or similarity in CDP literature. This chapter thus fills an important gap. Nevertheless, the method has been employed in a range of other political science and legal research . (Wilkerson & Casas, 2017). Alschner and Skougarevskiy apply the technique to bilateral investment treaties (2016). Spirling analyzes treaty-making between indigenous peoples and the U.S. Government, making the argument that the increasing military and economic power of the U.S. resulted in worsening terms for indigenous groups over the nineteenth century (Spirling, 2012). Legal-text-as-data has also been employed to evaluate the success of legal harmonization projects within the EU, (Sulis et al., 2021), and the rate of reuse in bill text in American legislative text, (Linder et al., 2020), (Burgess et al., 2016), as well as the changes in focus of party manifestos in Western Democracies between 1960 and 2016 (Düpont & Rachuj, 2022). In short, textual similarity can help explain differences across analytically significant groups, whether chronologically organized or organized by political valence. Such differences might reflect changes in bargaining or public choice dynamics, policy innovation, legal paradigm shifts, or as here, processes of convergence. (Alschner & Skougarevskiy, 2016, p. 34). To be sure, textual similarity analysis goes beyond group comparison. It also presents a ”tractable resource” with which to trace diffusion of text across jurisdictional contexts. (Shaffer & Elkins, 2019, p. 4). Wilkerson et al. (2015), for example, employ it to trace the flow of ideas across American states by identifying instances of progressive text reuse in legislation. Similarity can also reveal changes in language over time or continuations in language in response to an external event. Garrett and Jansa (2015), for instance, show how American states vary in the degree to which they adopt legislative formulations promulgated as model legislation by interest groups.

Treating legislative texts as data does have downsides. (Grimmer & Stewart, 2013). “Automated text comparison works especially well when the task involves comparing text rather than interpreting content.” (Alschner, 2020, p. 8). Any conversion of legal or political documents to data leads to a loss of semantic information. (Alschner, 2020, p. 3). Human readers understand subtly, tone, and variation in diction. They bring with them their expertise and their understanding that not every word is important to the meaning. Text-as-data cannot evaluate such elements. Text-as-data is also frequently blind to small changes that create large changes in meaning. Consider the sentences “he did run” and “he didn’t

run." To a human reader these are exact opposite sentences, but because they only vary by three characters, simplistic text-as-data approaches may treat the two as highly similar. Moreover, text will often fail to capture "tacit knowledge" or peripheral information known to the parties. (Moodley et al., 2019, p. 70). All "text[s] [are] constructed through information outside of the text." (Alschner, 2020, p. 3). Legislative text that drafters and relevant officials do not intend to enforce, for example, is meaningless regardless of its textual features. And again, text-as-data is blind to these elements. These downsides can be mitigated somewhat with careful qualitative analysis that compares the statistically similar texts for semantic meaning, as in Part II.

§3.2 Methods

This section describes the data collection process, then addresses the resulting dataset, the methods employed to clean the dataset, and finally the similarity measures and statistical software used to evaluate the resulting plain text corpus.

§3.2.1 Data Collection

For this project, approximately 260 countries or subnational jurisdictions, including American states, Canadian provinces, several overseas territories of the U.K., and the Faroe Islands, were included in the scope of analysis. Of those approximately 260 jurisdictions, 210 have a data protection law presently in force according to the definition of data protection laid out in the Introduction. Thus, there were 210 possible first adoptions of data protection law to collect. Using a variety of sources, 203 PDFs of the first enacted data protection legislation were collected. No copies of the Faroe Islands' 2001 data protection law, San Marino's 1983 data protection law, the Democratic Republic of the Congo's 2020 data protection law, or Niger's 2017 data protection law were available online, though the author was able to confirm that each exists with independent secondary sources. Samoa's 2005 Telecommunications Act, which is sometimes treated as a data protection law, appeared to be only tangentially related to data protection goals, particularly personal privacy (which it mentions only once in 74 pages of rules for telecommunications companies) and so was excluded. The 2008 Belarussian data protection law was located, but it was stored in a blocked Belarussian government database. And finally, the legitimacy of a PDF of Chad's 2015 data protection posted online could not be verified with independent sources.

The remaining 203 laws were sourced from a variety of online databases. Chiefly, PDF copies of legislative texts were sourced from the National Data Privacy Legislation database maintained by the World Legal Information Institute ("WorldLII"). WorldLII is a free-to-access single search facility for legislation and legal code maintained by the Australasian Legal Information Institute. PDFs for approximately 40% of the total corpus were obtained from sources outside the National Data Privacy

Legislation database, particularly from the OneTrust DataGuidance database, which OneTrust maintains on a per country basis by contracting out to local attorneys. The remaining PDFs (approximately 5%) were sourced from targeted searching in government databases. In a small minority of cases, text was manually copied from a government website into a .docx document and then transformed into a PDF. That introduces a small but non-trivial possibility of typographic errors. One law, Uruguay's 2008 *Ley de Protección de Datos Personales*, was excluded at the collection stage because the available PDF file was too occluded to convert to text either manually or using Optical Character Recognition.

Once laws were collected, they were machine translated to English where necessary. Only languages in which the author could spot check the accuracy of the translations using a legal dictionary in the source language were included in the final dataset.[17] Translations were accomplished using DeepL or Google Translate's API. Source PDFs with multiple columns of different languages side-by-side or occluded PDFs did not translate properly using available translation software. Thus, Burkina Faso's 2004 legislation was excluded as was Morocco's 2009 law, Algeria's 2018 law, Angola's 2011 law, Aruba's 2011 law, and the Republic of Congo's 2019 law, among several others. Multiple columns of foreign language legislative text was a particular issue in North and West Africa.

After translation to English and verification of the translation, PDFs of a first data protection law in 143 countries and subnational jurisdictions remained. Finally, the GDPR as well as the agreements and principles enshrined in Convention 108 and the 1980 OECD Guidelines were added to the set. Even though the United States is typically not included in the list of countries that have a data protection law, a PDF of the legislative text of the 1974 Privacy Act was included as an example of a first-generation data protection law for public entities and agencies. The Uniform Law Commission's Uniform Data Protection Law template was also included in the 2021-2023 bracket to determine whether it had been widely copied in American states (it has not).

§3.2.2 Dataset

The included subset of countries and subnational jurisdictions represents 71.9% of the first adopted legislative texts in existence as of September 2024. The dataset spans the globe and is regionally diverse. (See Fig. 3.1). The countries represented are also diverse in size and trade volume, ranging from St. Lucia and Malta to India and China. The corpus also reflects an accurate picture of the

[17] This meant that languages such as Lao, Russian, Croatian, and Mongolian were excluded. Languages were also excluded if they were the only instance of that language in the collected PDFs such that comparison across translations for the reliability of translation software (i.e., that it returned the same results for the same words across PDFs) was not possible. Thus, Indonesia's 2013 legislation, which was the only legislation in Indonesian, was excluded. Azerbaijan's 1998 law was excluded, post translation, because translation revealed that only excerpts were available rather than the full legislative text.

spread of data protection rules over time. It includes 10 rules from 1973-1980, 9 rules from 1981-1990, 19 rules from 1991-2000, 31 rules from 2001-2010, 40 rules from 2011-2020, 14 rules from 2021- March 2023, and 25 rules adopted between March 2023 and September 2024.[18]



Figure 3.1. Included countries with sub-national jurisdictions of the U.S. included in inset. White = not included. Grey = included. Black = subnational jurisdictions included.

§3.2.3 Cleaning the Data

As Linder et al. note, "legislation is full of textual content that is not very useful in identifying" meaningful similarity. (Linder et al., 2020, p. 556). Once the PDF files were collected, meaningless content was removed from the raw PDFs the data was cleaned using a separate workflow for PDFs that require manual cleaning as they did not store text data (e.g., were scans of legislation).[19] For PDFs that required manual cleaning, unnecessary pages of the PDFs were deleted. These included title pages, signature pages, and blank pages. The PDFs were then converted to .txt files using online PDF-to-text Optical Character Recognition software. The resulting .txt files were manually compared with the original PDF files to ensure no data loss. In some cases, information from the PDF files was manually typed into the .txt file where the PDF file was visually occluded, impeding automatic conversion. Symbols and typographic errors due to the conversion were removed as were schedules, annexes, tables of contents, and footnotes. Meaningless legal language, including boilerplate clauses inserted into all legislation in a

[18] The file size also varies considerably from Paraguay and the OECD, which are about 3 pages of text each at 6 kilobytes, to GDPR and Jamaica, each nearly 100 pages of text at more than 200 kilobytes each. Together the legislative texts make up an uncleaned text corpus of about 18.6 megabytes.

[19] Earlier texts (pre-1990) were significantly more likely to be manually cleaned since PDFs of early legislation were much more likely to be either occluded or a series of scanned pages without associated stored text.

specific country was also removed.[20] Sometimes typographic errors existed in the translated .txt file. These were corrected where possible using sentence-level context.[21] For PDFs that included stored text information, the R package pdftools was employed to convert the files to plain text. This means that most texts were not manually cleaned during the conversion.

Once the .txt file of each PDF was generated, the .txt files were preprocessed using the R text miner package ("tm"). A balance between simpler data and minimal information loss was sought. Since possible inferences from data are sensitive to preprocessing choices, (Denny & Spirling, 2018, p. 4), only those aspects of text data that were clearly irrelevant to the research question were excluded. As such, numbers, Unicode punctuation, and English-language stop words like "the" "which" "on" "at" and so on were removed from each .txt file. German-language stop words were also removed for purposes of Section 3.4.1. An author-selected list of common but semantically meaningless legal-legislative language was also removed from each file (see Table 3.2). Finally, each .txt file was transformed to lower case and white space was condensed. Each is a common preprocessing step for legal texts. (Moodley et al., 2019, p. 66); (Sulis et al., 2021, pp. 189–191).

Category	Words Removed
Stop words	(e.g.) "I" "at" "me" "my" "myself" "we" "our" "his" "himself" "it" "its" "the" "which" "on" "that" "the" "be" "been" "being" "is" "isn't" "was" "wasn't"
Selected words	"Article" "article" "Chapter" "chapter" "No" "no" "gazette" "Gazette" "whereas" "Whereas" "Seen" "seen" "Code" "code" "NDPrivLegis" "Art." "Sec." "SEC." "Section" "section" "Machine Translated by Google" "machine translated by google" "www" "unofficial translation" "Unofficial Translation"

Table 3.2. Removed words.

In total about 4.4 megabytes of data was removed from the plain text corpus, leaving a corpus of 14.2 megabytes. This corpus was then assessed for statistical similarity between each .txt file and all other .txt files in a single matrix as well as in a separate workflow for certain time periods. Each period was structured as its own matrix of similarity measures, where each entry in the matrix represented the distance between a pair of data protection texts (aka a dyad).

§3.2.4 Similarity Measures and R Environment

[20] An example being a formal execution phrase such as "Das vorstehende Gesetz wird hiermit ausgefertigt und wird im Bundesgesetzblatt verkündet," which was removed from the 1977 West German legislation.
[21] For example, the Section entitled "Right of Access" in Austria's 1978 Data Protection Act was originally translated into a confusing English tense so the tense was rewritten in the .txt file from surrounding context by the author.

There are a number of mathematical measures of similarity that can be employed to compare text (see Wang and Dong (2020) for an overview). In assessing similarity, two statistical measures common to the data science literature were used here: 5-gram Jaccard Distances and 5-gram Cosine Distances. Jaccard Distances and Cosine Distances measure how dissimilar or "distant" two texts are as a statistical matter of string distance.

5-gram refers to a contiguous sequence of 5 items—in this case words—in a given sample of text. In the sentence, "the dog ran fast over the large hill," a 5-gram sequence might be "the dog ran fast over" or "fast over the large hill" or "dog ran fast over the," and so on. 5-gram sequences were employed for two reasons. (1) There is some evidence that in English-language legal texts the validity of n-gram approaches compared to manual assessment tends to peak at n = 5. (Moodley et al., 2019, p. 66). This is possibly because shorter strings of words are overinclusive (a 3-gram, e.g., would tend to render all texts with the phrase "Article One Definitions" statistically similar), and long strings of exactly the same words are simply rare. 5-gram sequences capture the idea that copied legislative passages are likely to be longer than 5 words but not much longer. (2) While there is some evidence that similarity rankings for legal and political text are most stable at n-gram = 2-4 as measured by Kendall tau distance (Sharpe, 2017), a slightly higher n-gram would weight findings against high degrees of similarity. Thus, selection of n-gram = 5 is a manner of validating any results indicating large amounts of textual similarity in comparison to benchmarks.

A 5-gram Jaccard Distance calculation is a two-step process. First, the sets of text were vectorized[22] by representing them mathematically as a set of frequencies of the 5-gram sequences in the text, following Spirling (2012) and Alschner and Skougarevskiy (2016, p. 11). 5-gram sequences were employed for two reasons. (1) There is some evidence that in English-language legal texts the validity of n-gram approaches compared to manual assessment tends to peak at n = 5. (Moodley et al., 2019, p. 66). This is possibly because shorter strings of words are overinclusive, (a 3-gram, e.g., would tend to render all texts with the phrase "Article One Definitions" statistically similar), and long strings of exactly the same words are simply rare. Put differently, 5-gram sequences capture the idea that copied legislative passages are likely to be longer than 5 words but not much longer. (2) While there is some evidence that similarity rankings for legal and political text are most stable at n-gram = 2-4 as measured by Kendall tau distance (Sharpe, 2017), a slightly higher n-gram would weight findings against high

[22] Sometimes the NLP literature refers to this step as dimensionality-reduction.

degrees of similarity. Thus, selection of n-gram = 5 is a manner of validating any results indicating large amounts of textual similarity in comparison to benchmarks.

Once vectorization was complete, the Jaccard distance between the vectorized texts was computed. The Jaccard Distance between vectors is specified by 1 minus the 5-grams present in both sets over all 5-gram sequences in either set. In set notation this is represented as $1 - |X \cap Y|/|X \cup Y|$ where X is the first text and Y the second. Results were arrayed in a symmetrical matrix, with each value representing a dyad's Jaccard Distance.

The vectorization process was repeated for Cosine Distance. The Cosine Distance between the vectors is specified by 1 minus the Euclidean dot product of vector A, where A is the term frequency vectorization of the first text, and vector B, where B is the term frequency vectorization of the second text, over the magnitude of each vector. Mathematically, this is given by the equation: $1 - \frac{AB}{||A||B||}$. The basic intuition behind Cosine Distance is that the more similar the two vectors are, the smaller the angle between them will be, and thus the smaller the Cosine Distance. Results were again arrayed in a symmetrical matrix.

Both Jaccard and Cosine Distances have strengths and weaknesses. Cosine Distance is usually independent of the length of texts because it controls for the magnitude of vectors. Jaccard Distance, on the other hand, does not perform well compared to human analysis in situations where a long text is compared to a short text (e.g., GDPR to Paraguay's data protection law) since the short text will have far fewer n-grams, meaning the amount of shared n-grams is low regardless of text reuse or copying. Still, Jaccard Distance, because it relies on the n-grams present in both sets over the 5 grams present in either set, avoids inflating the similarity of short texts to one another, an occasional issue with Cosine Distance. (Shaffer & Elkins, 2019, p. 15). Jaccard Distance, unlike Cosine Distance, also avoids inflated similarity measures in situations where word frequency is very high. (Zhou et al., 2022). This is likely to be an issue when comparing texts within a specific legal domain. In this project, for illustration, one would expect words like "privacy" "data" and "protection," to be repeated many times in legislative texts. Given these overlapping weaknesses both measures are employed herein but Jaccard Distance is relied on for most analysis, especially of dyads involving a relatively short text. Cosine Similarity is referred to as an overarching measure and when comparing short texts to long ones or vice versa.[23]

[23] The computations necessary to create matrices of Jaccard and Cosine distance were carried out in R version 4.3.0 for Windows. They were executed in an RStudio Environment with gplots, NLP, pdftools, readtext, stringdist, tm, Rfast, and base R along with the datasets, graphics, grDevices, stats, methods, and utils packages running.

§3.3 Contextualizing Similarity in Legislation and Law: Benchmarking

Unfortunately, statistical similarity is an inherently relative inquiry. To contextualize the textual similarity measures employed herein, this chapter utilizes a novel benchmarking approach. Certain laws are known copies of other laws. This is often due to copying from metropole to periphery in a colonial (or post-colonial) situation. A good example of known copying is the Indian Companies Act of 1956, which is a direct, almost clause-for-clause copy of the UK 1948 Companies Act. (Spamann, 2009, p. 1813).

Certain laws are also known to be heavily influenced, but not direct copies, of an earlier law. For example, the American 1790 Copyright Act is known to be based on the 1710 Statute of Anne in the U.K. but has its own specific meanings and constructions. (Bracha, 2010, p. 1473). Another good example of law that was strongly inspired by but diverged in important respects from a set of rules in a separate jurisdiction is Singapore's 2001 "U.K.-inspired" Code of Corporate Governance. (Puchniak & Luh Luh, 2017, p. 267). A further example is the Italian Criminal Code of 1988, which imported some aspects of American criminal procedure, including its hearsay rules, but modified those rules in both text and meaning to suit distinctly Italian civil law preferences. (Grande, 2000, p. 247; Panzavolta, 2016). Such laws, because the degree of influence between clause-for-clause copying and no copying at all is broad should exhibit a wider range of statistical Jaccard and Cosine Distances than direct copies. They should therefore provide a range of influence, from strongly inspired and practically copied (Singapore and UK Codes of Corporate Governance) to loosely inspired and translated only in spirit rather than copied per se (e.g., Italian, and American hearsay rules) with which to assess data protection laws.

At the other end of the spectrum from direct clause-for-clause copies are laws that have no theoretical reason to be copies of one another even though they pertain to the same legal topic. A good example might be the definitions of public benefit that underlie certain exceptions from tax rules in various states. *Prima facie*, one might expect such definitions to be deeply linked to culture, and to what purposes a given society finds especially valuable whether educational, collective, charitable, or political. And since there is no indication that the definitions of public benefit in the legal codes of Germany and the United States are linked to each other through the sharing or borrowing of ideas, they might be taken as an example of distinct rules.

Using these examples, a table of benchmarks can be created to compare statistical distances in the data protection space (Table 4.1). As expected, known copies have smaller statistical distances (i.e., be more statistically similar) than do instances of known influence and distinct legal rules.

Benchmark	Rules	Jaccard Distance	Cosine Distance
Identical Rules	Statute of Anne 1710 UK Statute of Anne 1710 UK	0	0
Known Copying	India 1956 Companies Act UK 1948 Companies Act	.541	.147
Known Influence Range	Singapore Code of Corporate Governance 2001 U.K. Code of Corporate Governance 2000	.632	.162
	Copyright Act of 1790 US Statute of Anne 1710 UK	.768	.420
	Art. 195 Codice di procedura penale 1988 Federal Rules of Evidence Rules 801-804 (1988 version)	.921	.710
Theoretically Distinct and Unrelated	IRC §501(c)(3) U.S. Sections 52-54 of the Fiscal Code Germany (*Abgabenordnung*)	.926	.795
Completely Distinct Rules (no shared n-grams)	String "abc" String "def"	1	1

Table 3.1. Benchmarks.

Based on the skeleton story of convergence presented in Chapter 2, and the benchmark table above, a set of basic hypotheses and expectations might be stated. First, one would expect there to be an average Jaccard and Cosine Distance indicative of at least strong influence if not outright copying in the corpus (< .5 and .15 respectively). Second, over time the degree of similarity observed should increase, and both distance measures should decrease, likely exponentially year-on-year or decade-on-decade.

§3.4 Results

This section begins with a series of robustness checks. It then proceeds to a general overview of the entire corpus before sub-setting the corpus by year to visualize changes over time. While the full

corpus results clearly show degrees of similarity consonant with known influence on legislation, the subset results indicate a heretofore underappreciated trend in comparative data protection: the grouping of laws as mini-convergences.

§3.4.1 Robustness Checks: Machine Translation and First-Passed as Proxy

The internal validity of the results below depends on two assumptions: that machine translation neither introduces large degrees of dissimilarity/similarity nor irregularly distributed degrees of dissimilarity/similarity and that the first-passed legislation is a reasonable proxy for a data protection regime overall. This subsection provides evidence for both assumptions using Germany as a test-bed. Results show that both assumptions are justified.

In several ways Germany is the perfect test for both assumption one and assumption two because the test is stacked against finding evidence for both assumptions. Regarding assumption one, there are many German data protection texts all written within the same legal system increasing the possible sample size for translation purposes (see Table 3.3). Moreover, the linguistic variety of legal German compared to legal English suggests that one should expect introduction of textual similarity not present in the native German texts via machine translation. One would also theoretically expect major changes to data protection law after the 1983 Census Decision and German reunification in 1990 which introduced Lander with previous experiences of non-democratic, surveillance-heavy governance.

State	Year of First-Passed Legislation	GDR or FRG
Baden-Württemberg	1979	FRG
Bayern	1978	FRG
Berlin (West)	1978	FRG
Brandenburg	1992	GDR
Bremen	1977	FRG
Hamburg	1981	FRG
Hessen	1970	FRG
Mecklenburg-Vorpommern	1992	GDR
Niedersachsen	1978	FRG
Nordrhein-Westfalen	1978	FRG
Rheinland-Pfalz	1974	FRG
Saarland	1978	FRG
Sachsen	1991	GDR
Sachsen-Anhalt	1992	GDR
Schleswig-Holstein	1978	FRG
Thüringen	1991	GDR

Table 3.3. Landesdatenschutzgesetz.

Of the German state data protection laws (*Landesdatenschutzgesetz*) outlined in Table 3.3, PDFs of the first-passed law in Schleswig-Holstein, Mecklenburg-Vorpommern, Sachsen, Hesse, and Rheinland-Pfalz were collected. These laws, in their German version, were then compared using Jaccard Distances against the German language versions presently in force in each state as well as the 1977 Federal

Republic of Germany ("FRG") legislation and contemporary German-language versions in Bavaria, Saarland, and Saxony-Anhalt. The same procedure was performed for the machine-translated English-language versions of each legislative text. As one can see from Figures 3.2 and 3.3, the results are visually similar. One can observe a block of texts passed in 2017 and 18 in the center of both heatmaps as well as a cluster in the bottom left corner of 1970s legislation and, in the upper right corner, the first-passed and contemporary texts in Saxony. Visual analysis is confirmed by basic correlation analysis and descriptive statistics. Overall, the average Jaccard Distance of the German-language matrix is .671 while the average for the English-language version is .635. As expected, machine translation introduced a small degree of similarity to the corpus. Still, the amount is minimal and distributed relatively evenly across the dyads. In other words, no dyad's Jaccard Distance changed more than any other dyad's distance.[24] This is reflected in basic correlation testing between the matrices as strings of numerical values which returns a correlation of .993, indicating very high correlation between the matrix values pre- and post-translation.

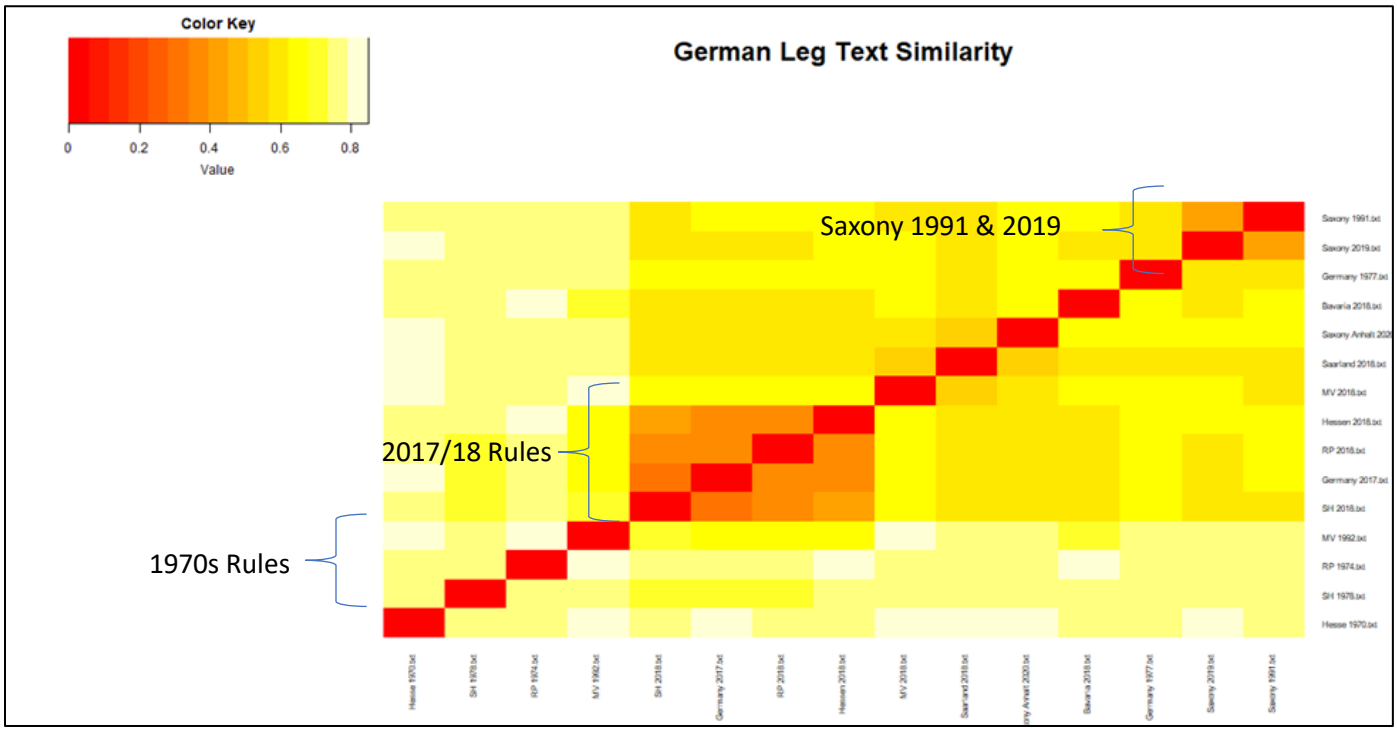


Figure 3.2. Heatmap of textual similarity of German rules in German measured by Jaccard Distance. Colors towards the red end of the spectrum indicate lower Jaccard Distance values, and higher degrees of similarity.

At the same time, the Jaccard Distances between the earliest and latest passed rules in Germany at both the national and subnational level indicate that while rules have changed since the 1970s, they

[24] Bavaria's values appear most affected by machine translation.

have not changed so much as to abandon any textual constructions from early rules. The German-language Jaccard Distance between the earliest rule in Hesse and the most recent rule in Saxony-Anhalt is .8, a value approaching the mid-range of known influence benchmarks. The highest distance overall, .847, between Rheinland-Pfalz's 1974 rules and Mecklenburg-Vorpommern's 1992 rules, is also indicative of known influence. It is also apparent that some German jurisdictions have done very little updating since their first-passed rule. In fact, Germany's 1977 national-level legislation has a range of Jaccard Distance between .64 and .79 across dyads, distances comparable to a situation of known legislative influence. And Saxony's 1991 and 2019 rules are textually more similar (.404) than all but two jurisdictions in the full corpus dataset (see Appendix A).

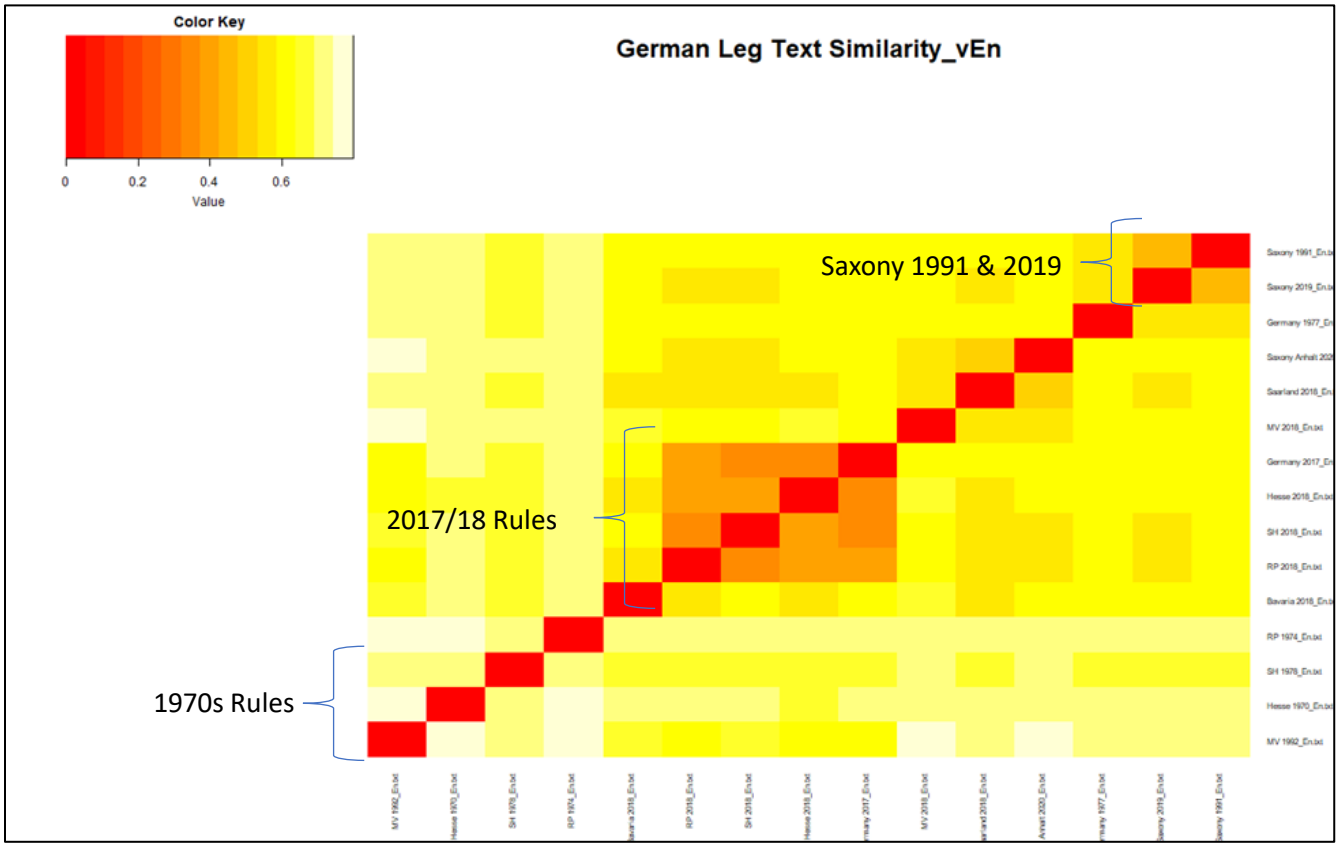


Figure 3.3. Heatmap of textual similarity of German rules in English measured by Jaccard Distance. Colors towards the red end of the spectrum indicate lower Jaccard Distance values, and higher degrees of similarity.

So, while it remains true that German jurisdictions have been updating their rules over time—change should be expected—they have not changed so much as to render the first passed rule, particularly at the national level in 1977, a poor proxy for German texts across time. In sum then, the German case provides evidence both that machine translation does not affect results, except to introduce marginal amounts of similarity compared to standard German (*Standardhochdeutsch*), and that the first-passed legislation is a valuable proxy for all legislative texts in a selected jurisdiction.

§3.4.2 Full Corpus Results

In general, the matrix results of the Cosine and Jaccard Distances indicate high degrees of textual similarity compared to benchmarks. The average Cosine Distance of texts in the corpus is .386, a Cosine Distance value consonant with known instances of legislative influence. The average Jaccard Distance of texts is .752, again consonant with known instances of legislative influence. These averages, while informative, obscure radically left and right skewed distributions respectively (see Fig. 3.4). Jaccard Distances are highly concentrated between .648 and .864. This is roughly equivalent to the benchmark range of known influence. Cosine values are highly concentrated between .1 and .3, again indicative of known influence between legislative text dyads.

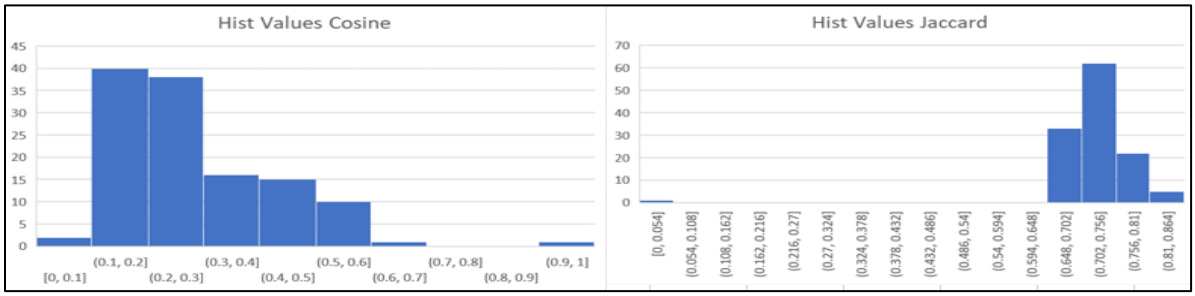


Figure 3.4. Histogram of Cosine Distance values in left panel. Jaccard Distance values in right panel. Generated in Excel from source matrix.

Counter to expectations, as Figure 3.5 shows, yearly average similarity distances, both Jaccard and Cosine, across dyads in each year do not appear to decline significantly over time, even if all show known influence compared to benchmarks. If anything, the range of distances increases over time with earlier years tending to have a smaller range and lower top-end similarity value. This may be because unique laws that do not show similarity distances consistent with known influence seem to appear more frequently after about 1995, which could cause fluctuation in yearly averages.

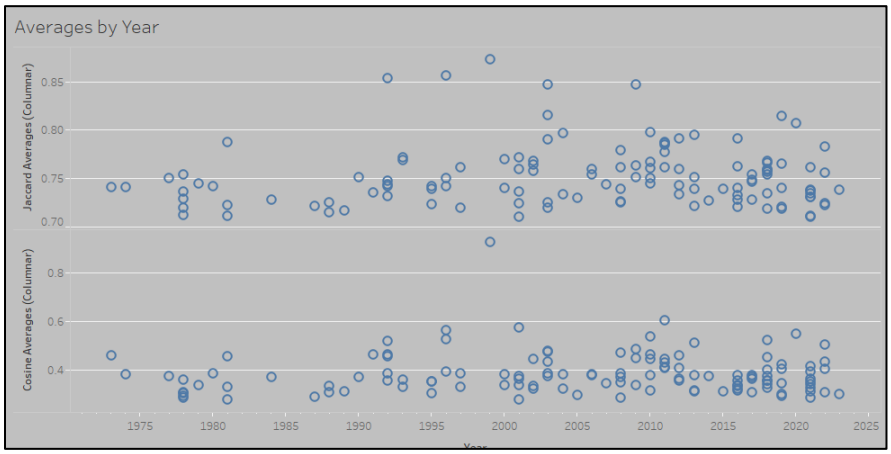


Figure 3.5. Top plot = average Jaccard Distance by year (range between .7 and 1). Bottom plot = average Cosine Distance by year (range between 0 and 1).

Still, the pattern is weak. As Table 3.4 shows, year of adoption and yearly average similarity measure are negligibly, but positively, correlated (see Table 3.4). As the year of adoption increases, the average similarity distances tend to grow as well, meaning laws are becoming more dissimilar but at a vanishingly small rate. However, the correlations are not statistically significant. Instead, they may be related to the small per-year sample size, leading to fluctuation in the average per-year value. The expected pattern returns for Jaccard time-bucketed results. As expected, average Jaccard distances decline, roughly linearly, between time brackets from .810 to .778.

	Cosine	Jaccard
Pearson's R	.0329 (.718)	.0415 (.545)
Kendall Tau (non-parametric)	.0123 (.842)	.0374 (.649)

Table 3.4. Correlation of Similarity Measures and Year of Adoption. p-values in brackets.

What is clearer is that the percentage of dyads below benchmark similarity values indicative of influence and copying increase over time between time brackets across the corpus before dropping around 2020. Figure 3.6 shows that a large body of dyads with distances indicative of influence, if not outright copying, appears in 2000, and then over the course of the 2000s, in comparison to conservative benchmarks, a few dyads appear that are nearly exact copies of each other appear. The weak patterns in per year averages but drastic increase in dyads which show influence and/or direct copying calls for more detailed heatmapping of each time bracket to identify dyads that stand out as likely examples of influence or copying.

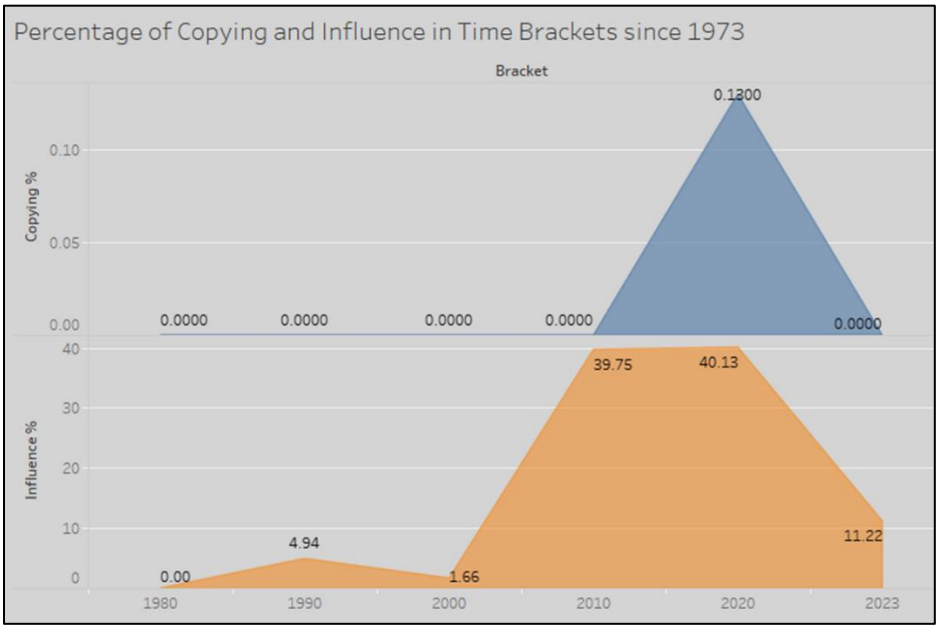


Figure 3.6. Percentage of rules in each time bracket that are at or below conservative benchmarks for copying and influence. Top = % of rules that show copying. Bottom = % of rules that show influence. Scales different between top and bottom.

Figure 3.7 presents a heatmap of the Cosine Similarity of every legislative text in the corpus to every other text in the corpus. Colors toward the red end of the spectrum indicate lower Cosine Distances and higher degrees of textual similarity. The bifurcation of legislative texts into two groups is clearly discernible, both visually and at the highest level of the columnar dendrogram. In the lower lefthand corner, early nations, nations adopting registration systems, and the recent U.S.-state level adoptions, apart from California, are grouped together with high levels of Cosine similarity between their legislative texts. The lower left-hand group also includes many developing nations, possibly suggesting that developing nation rules are more similar to one another than to the EU style rules that they are thought to emulate. In the upper right-hand corner are EU-region rules, particularly the DPD and rules like Poland's 1997 legislation, which was adopted to implement the DPD in Poland. There is a clear visual divide between the EU rules and those in the lower left-hand corner. At the same time, the generally low degrees of Cosine Distance (< .05) obscure potential patterns within the blocks. And these blocks of very low levels of Cosine Distance are indicative of high numbers of repeat terms and n-grams, which as noted above, may inflate Cosine Similarity.

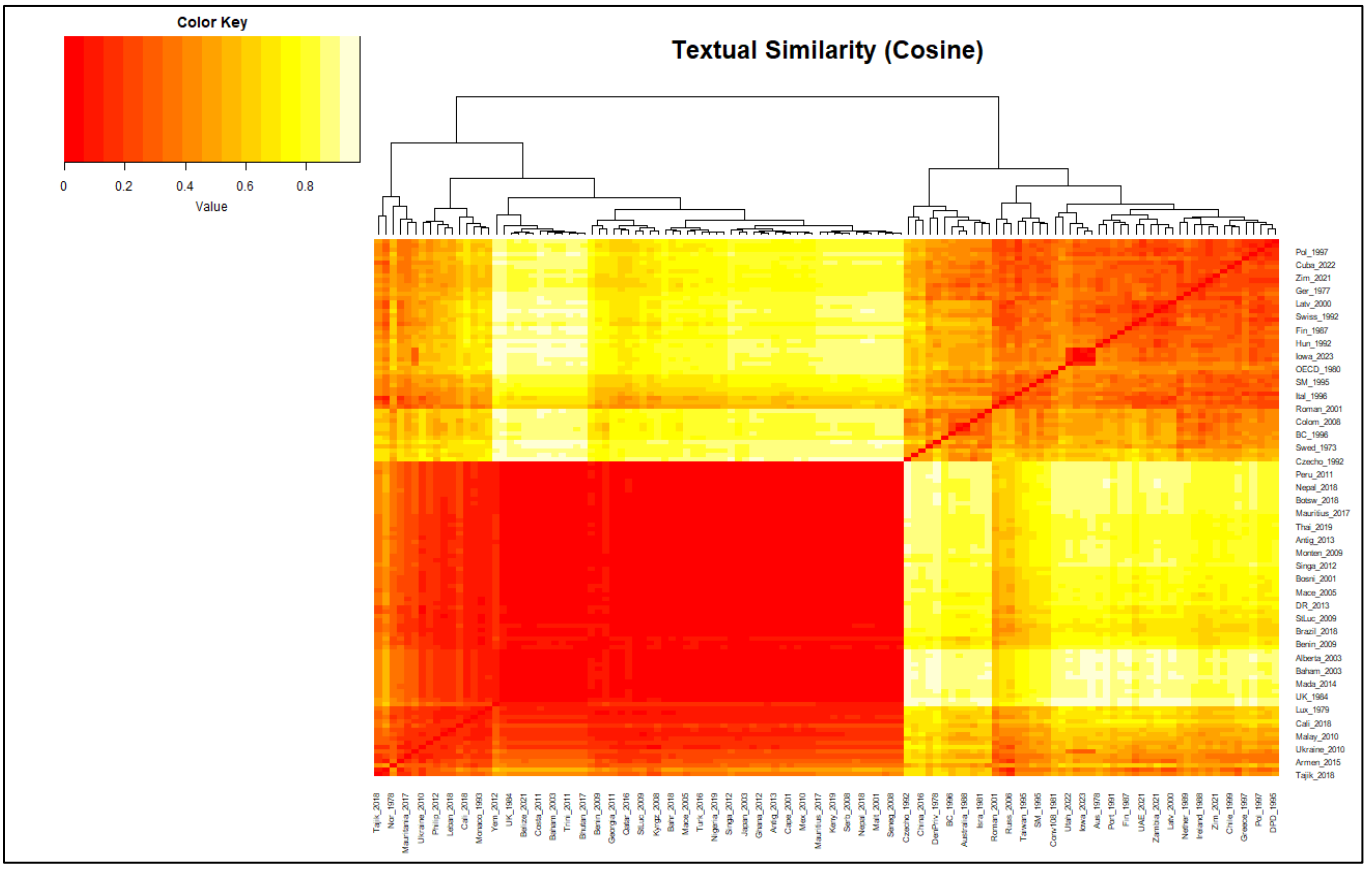


Figure 3.7. Heatmap of textual similarity measured by Cosine Distance. Dendrogram included for column measures. Colors towards the red end of the spectrum indicate lower cosine distance values, and higher degrees of similarity. Hierarchical sorting of axes activated. Data set includes texts through March 2023.

Figure 3.8 offers a Jaccard Distance heatmap of the same corpus. Since Jaccard Distance controls for shared n-grams, it offers a more nuanced picture of degrees of similarity than does Cosine Distance in this use case. The lighter colors in Figure 3.8 tend to indicate higher overall distances, meaning points of shared language stand out visually. In the case of Jaccard Distance, three groups are visible on the columnar dendrogram as opposed to two. First, on the far left are rules that are unique. These include Sweden's 1973 rules, which are known to have been influential but not widely copied, (Flaherty, 1989). Other jurisdictions in this group, including Yemen and Tajikistan, may have similarly idiosyncratic rules. Jurisdictions in the middle grouping roughly cluster around an American state grouping, an Island jurisdiction grouping, and in the upper right-hand corner an EU-style rules grouping, which includes recent adopters like Turkey, Bermuda, and the Philippines as well as earlier EU-region adopters like Monaco and Serbia.

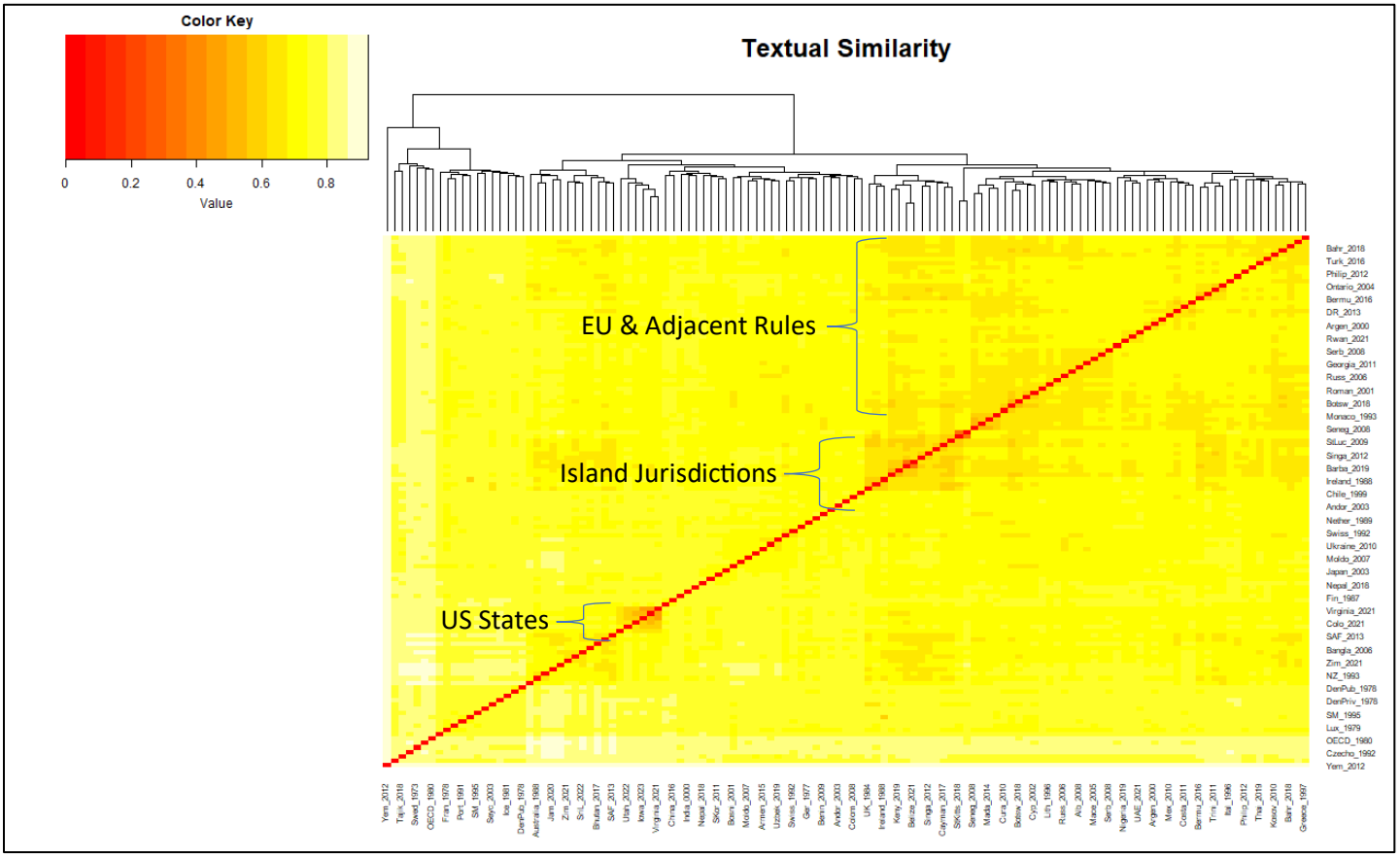


Figure 3.8. Heatmap of textual similarity measured by Jaccard Distance. Dendrogram included for column measures. Colors towards the red end of the spectrum indicate lower Jaccard Distance values, and higher degrees of similarity. Hierarchical sorting of axes activated. Data set includes texts through March 2023.

The Island jurisdiction grouping includes St. Lucia, Barbados, and Antigua and Barbuda. These three jurisdictions' first-passed legislative texts are more textually like one another than other dyads in the set. The

same is true for American states, particularly Iowa, Connecticut, and Virginia. Full corpus analysis also reveals some unexpected dyads. By both Jaccard and Cosine Distance, Belize and Barbados have the most similar rules.[25] Given these results, some expected and some unexpected, more fine-grained sub-setting of analysis is called for.

§3.4.3 Time-Bracketed Results

This subsection addresses time-bracketed results in seven categories. First, it heatmaps rules adopted between 1973 and 1980. This group includes most pioneer data protection laws. Second, it heatmaps legislation passed between 1981 and 1990. This second generation of rules coincides with the development of the convergence thesis in the comparative data protection literature. Third, it assesses rules passed between 1991 and 2000 before proceeding to the next two decades, 2001-2010 and 2011-2020, and finally visualizing the textual similarity of recently adopted legislation passed between 2021 and March 2023 as well as from March 2023 to September 2024. Since each time-bracket includes fewer laws and all laws in each bracket were passed within a similar period, one would expect changes in convergence dynamics over time to show up more clearly than in whole corpus analysis.[26]

Several overarching results can be discerned from the time-bracketed heatmapping. First, as might be expected, jurisdictions with large domestic markets are outliers in their respective time periods. These include India, China, and California, all of which are less textually similar to contemporaneous rules than other rules in the same time bracket. Second, convergence, as measured by a general shift toward the red end of the spectrum, is a clear mapping result over time. While the average Cosine and Jaccard Distances decline only slightly over time from 1973, the number of dyads with distances indicative of copying increases between time brackets. But, beginning in the 2011-2020 bracket, that "red shift" retreats as a visual phenomenon in favor of grouping dynamics. In visual terms, this trend is indicated by the reappearance of lighter tones against which blocks of darker red stand out. So, while convergence is a broadly accurate description, the emergence of grouping dynamics may be a more salient phenomenon since 2010. Groupings identified include jurisdictions that are geographically close to one another (e.g., East Africa or the Middle East) as well as jurisdictions that share a language (e.g., Costa Rica, the Dominican Republic, and Peru). Such groupings may represent mini or micro

[25] On face, there is a dearth of reasons to believe that Belize's 2021 law is an intentional copy of Barbados' 2019 rules. They could both be similar copies of other rules. This would be in keeping with the idea that many jurisdictions are copying the EU but the question is open.

[26] Section 3.4.3 employs Jaccard Distance as a measure to avoid the deep-red blocking exhibited by Cosine Distance in Figure 3.7. Hierarchical sorting was included to aid in visual analysis of year subsets that included more than 400 dyad pairs (i.e., heatmaps post-2000).

convergences. At the least, it seems descriptively accurate that groupings tend to represent countries with shared characteristics whether demographic, linguistic, or geographic.

Third, and most surprisingly, GDPR is among the least similar textually to contemporaneous rules. Unlike the DPD, there are no dyads indicating red-range, low Jaccard Distances between GDPR and legislative texts in other jurisdictions. If countries are copying European language, they appear to be copying a simplified, significantly parred down version of the legislation, rather than importing European legislation verbatim. Moreover, whereas the DPD was preceded by the introduction of divergent legislative texts in Europe (e.g., Slovenia's 1990 rules), GDPR was preceded by an increase in global grouping outside Europe (e.g., the South Africa or Madagascar-based groupings in Figure 3.13).

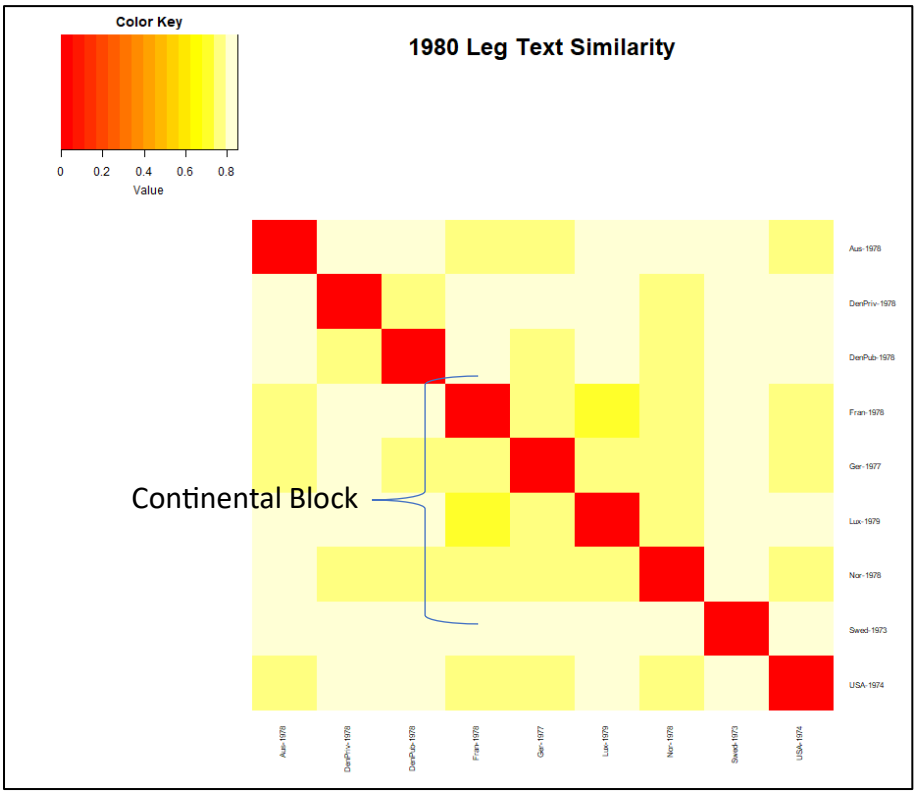


Figure 3.9. Heatmap of textual similarity measured by Jaccard Distance 1973-1980. Colors towards the red end of the spectrum indicate lower Jaccard Distance values, and higher degrees of similarity. No hierarchical sorting or dendrogram activated. US Privacy Act of 1974 included in bottom right.

Beginning with early data protection rules, Figure 3.9 shows relatively large Jaccard Distances (> .7) between the rules adopted between 1973 and 1980. These Jaccard Distances are consonant with known influence rather than with known copying. Sweden's 1973 legislation is clearly the most textually distinct as represented visually by the lighter colored bar second from bottom and second from right. Denmark's dual approach, two separate legislative instruments, one for the public and one for the

private sector, is also visually distinct. As one would expect, countries that share a language appear visually most similar. Thus, France and Luxembourg appear to have more overlapping n-grams than do other dyads. At the center of the heatmap, one can discern a continental European block with France, Luxembourg, and Germany. Scandinavian jurisdictions and the U.S. appear on the edges of this block.

Figure 3.10 shows the beginning of an overall red shift toward lower Jaccard Distances, and correspondingly higher degrees of similarity. This is consistent with the convergence literature. Later adopting states like Ireland (1988) and the Netherlands (1989) are the most similar dyads.

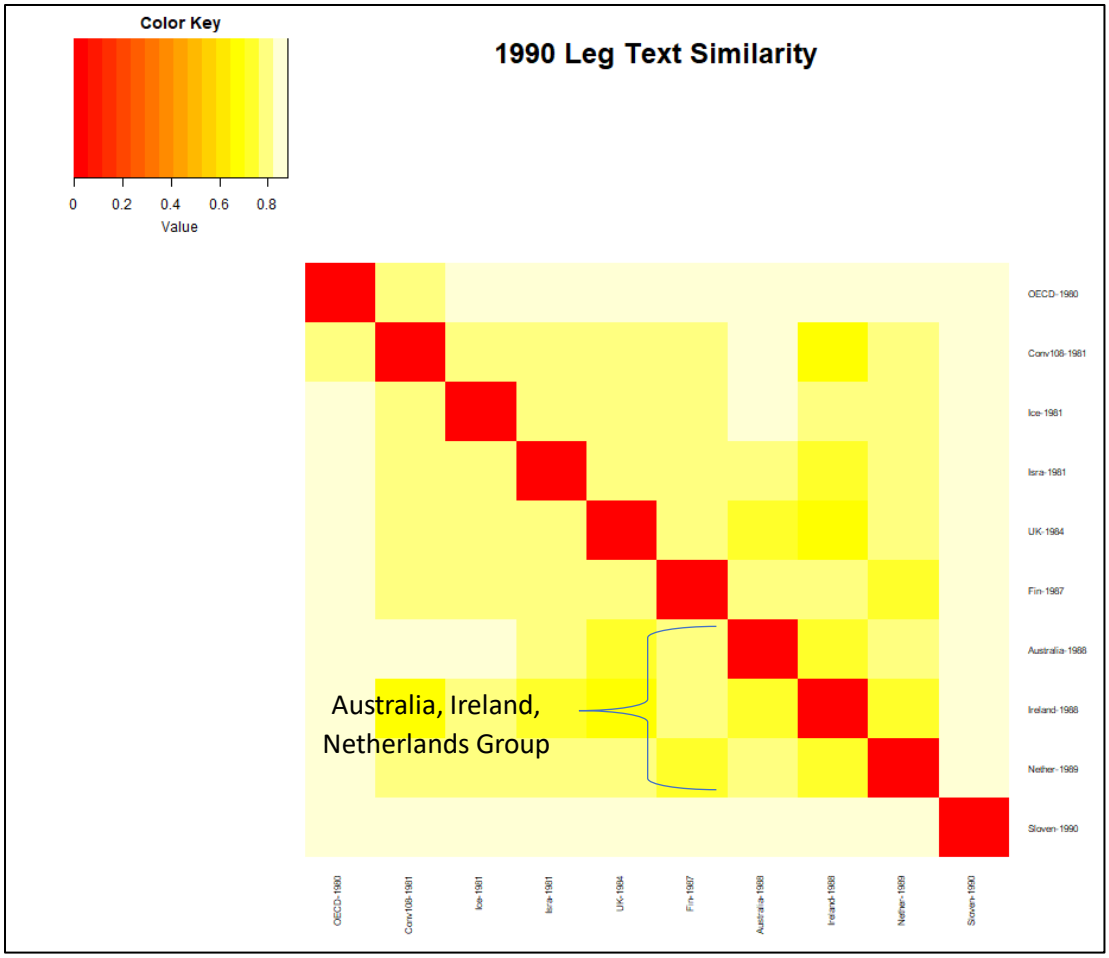


Figure 3.10. Heatmap of textual similarity measured by Jaccard Distance 1981-1990. Colors towards the red end of the spectrum indicate lower Jaccard Distance values, and higher degrees of similarity. No hierarchical sorting or dendrogram activated.

As a group, Australia (1988), Ireland (1988), and the Netherlands (1989) appear to indicate that in the 1980s, later adopting countries settled on a certain template developed over the course of the decade. At the earlier end of the decade, Convention 108 and the OECD Guidelines are most similar to each other as indicated by the box-like grouping in the upper left-hand corner of Figure 3.10. At the end of the decade however, there appears to be a break in the trend towards decreasing Jaccard Distance, as

Slovenia's 1990 rules, as indicated by the lighter bar in the right-most column and lowest row, is the most dissimilar across dyads.

Figure 3.11 continues the red shift. Lighter colors continue to become more infrequent. Outliers are clearly visible.[27] Czechoslovakia's rules are a point of dissimilarity. Poland, Greece, and the DPD are all very similar to one another, which makes some sense as Poland and Greece passed their first data protection legislation in the years following the enactment of the DPD in 1995.

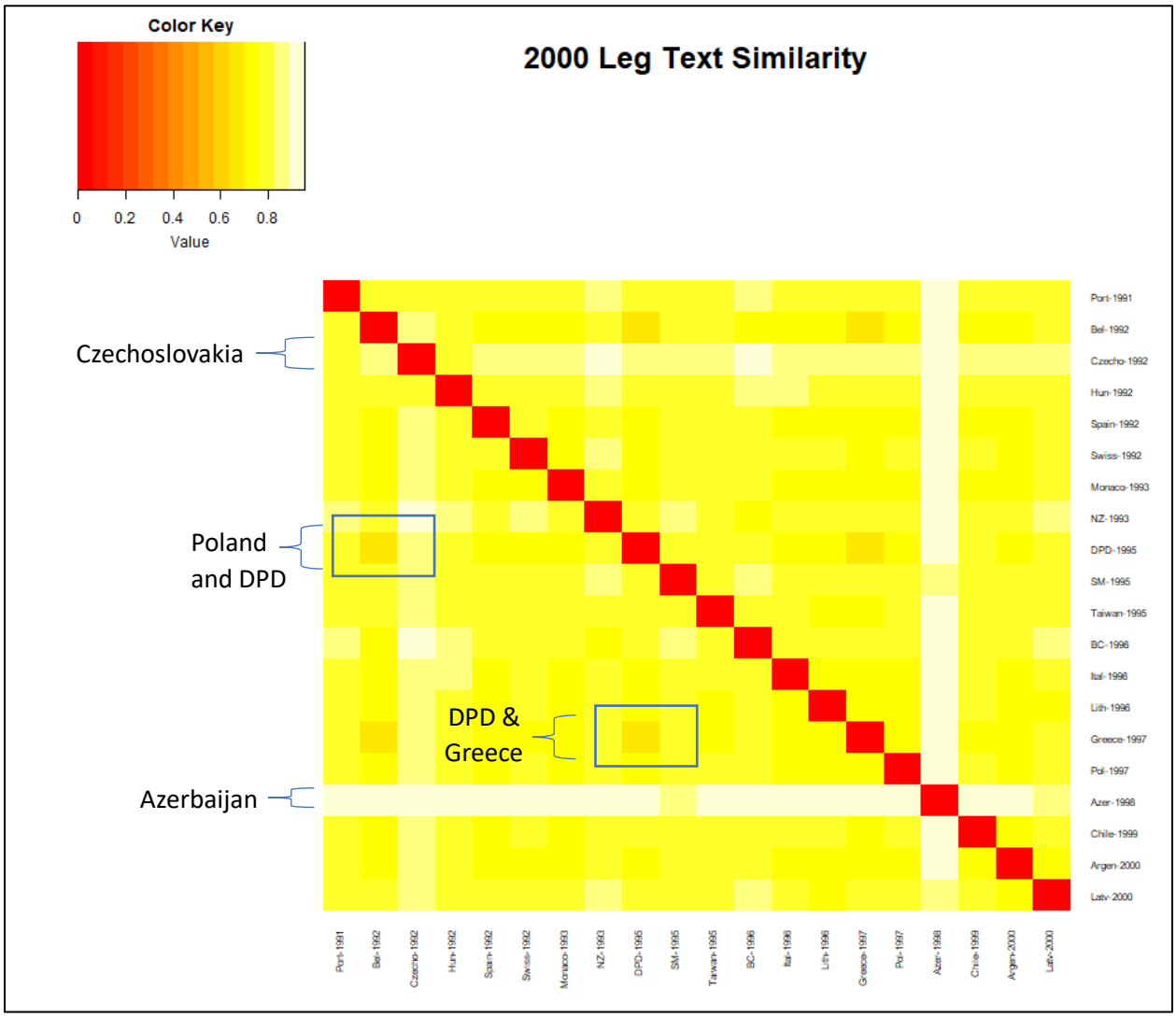


Figure 3.11. Heatmap of textual similarity measured by Jaccard Distance 1991-2000. Colors towards the red end of the spectrum indicate lower Jaccard Distance values, and higher degrees of similarity. No hierarchical sorting or dendrogram activated.

[27] These include Azerbaijan, which was included as a visual control; since it is only an excerpt, it should be the most dissimilar, and it is.

Figure 3.12 further continues the red shift, now into the orange range. Lighter yellows are even less frequent than in Figure 3.11. For the first time much clearer “blocks” or groups that have shared ranges of Jaccard Distance are visible. This includes two separate groups of island jurisdictions. In Island Group 1 the Bahamas and St. Lucia are very similar to one another. In Island Group 2, Cape Verde, Malta, Cyprus, and Curacao are grouped together in the orange range.

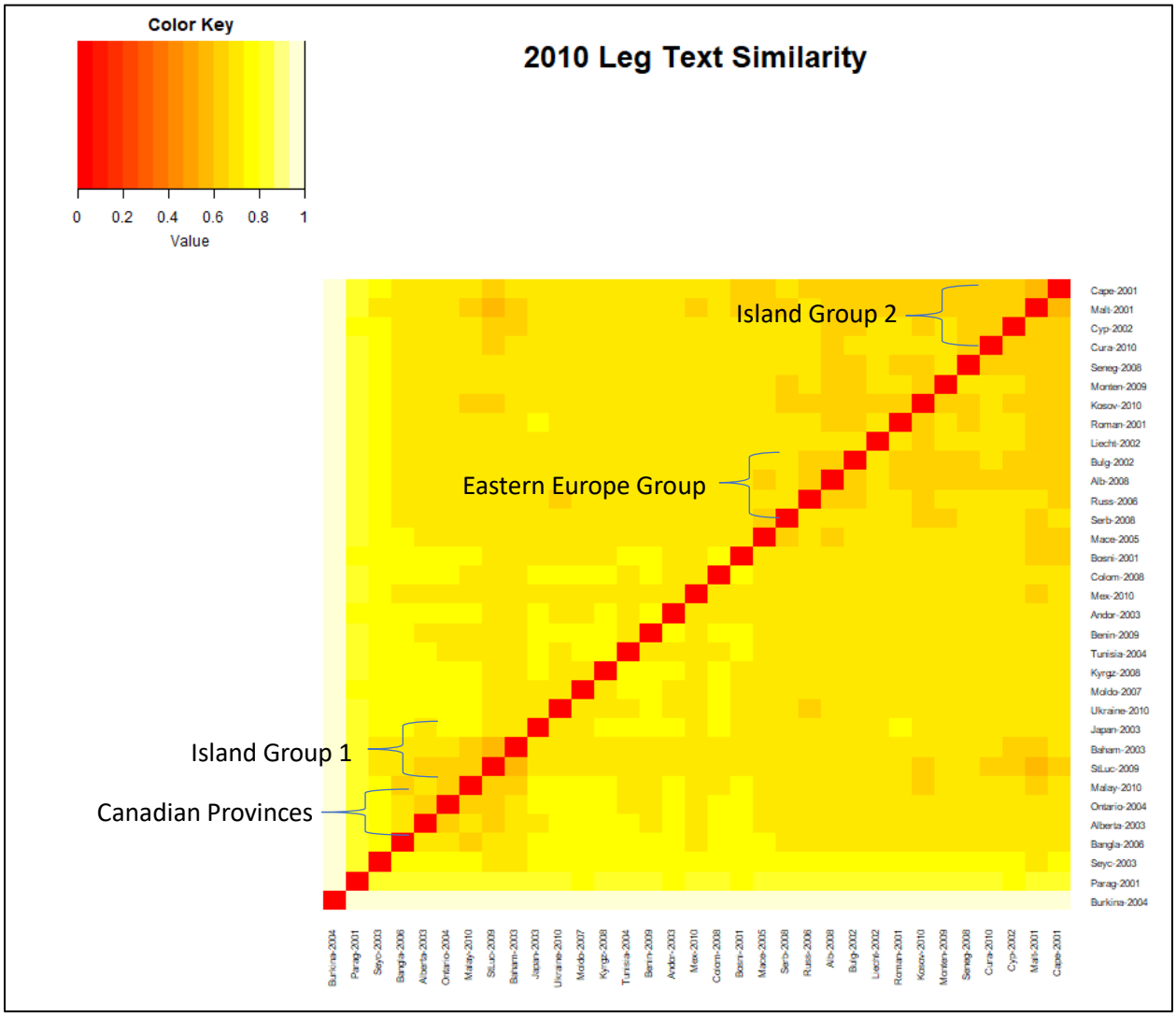


Figure 3.12. Heatmap of textual similarity measured by Jaccard Distance 2001-2010. Colors towards the red end of the spectrum indicate lower Jaccard Distance values, and higher degrees of similarity. Hierarchical sorting; no dendrogram activated.

There is also an orange-range block of Eastern European adopters including Albania, Bulgaria, Serbia, and Russia as well as a Canadian province block of Alberta and Ontario. Burkina Faso in the far leftward column and bottom-most row is a clear outlier, which makes sense. Its PDF file was retrospectively discovered to have been incorrectly translated by automated software, rendering its .txt file empty.

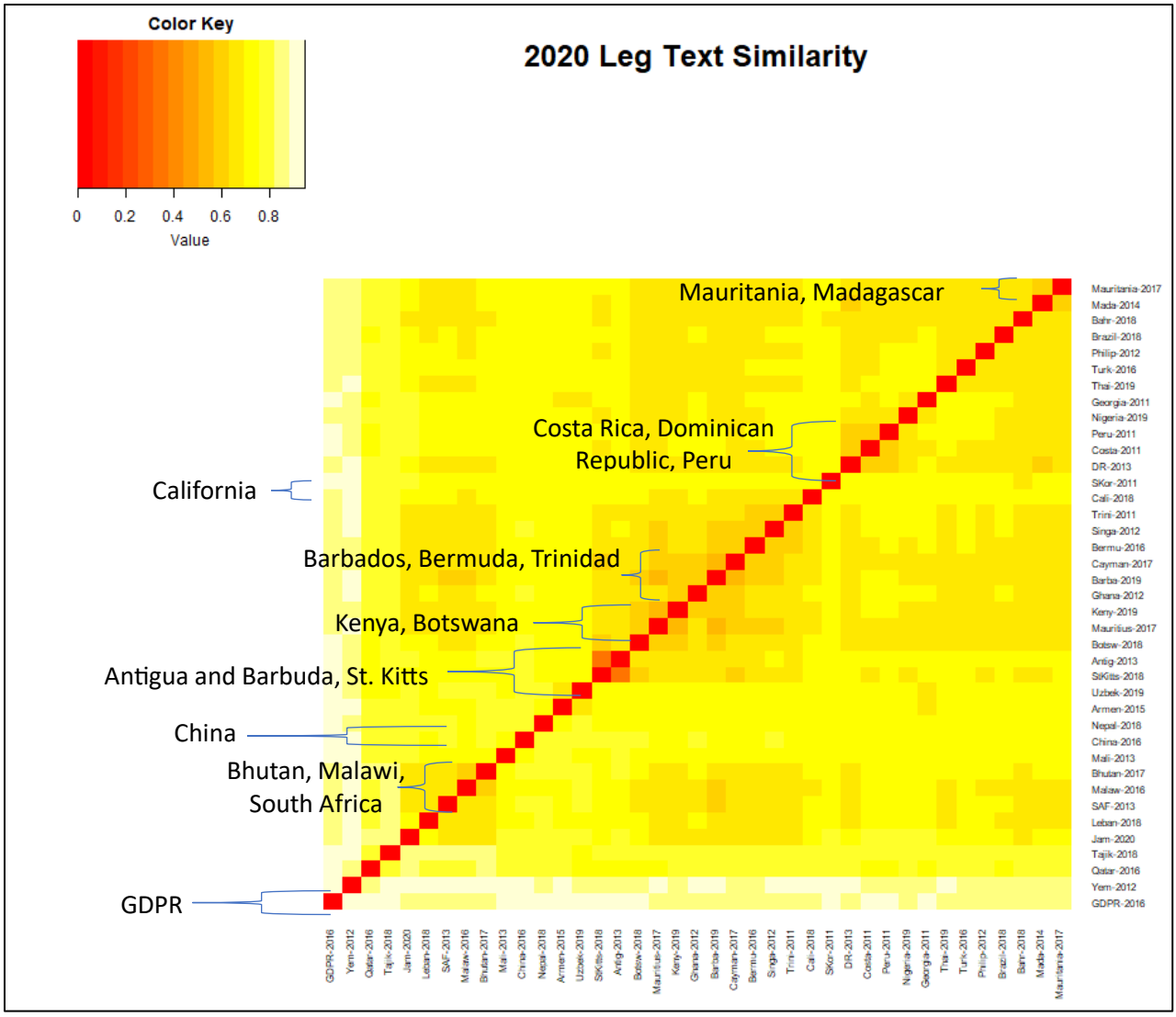


Figure 3.13. Heatmap of textual similarity measured by Jaccard Distance 2011-2020. Colors towards the red end of the spectrum indicate lower Jaccard Distance values, and higher degrees of similarity. Hierarchical sorting; no dendrogram activated.

Figure 3.13 continues the trend toward grouping even as it shifts back towards the yellow end of the spectrum with fewer orange-coded dyads than Figure 3.12. Dark red groups indicating very low Jaccard Distances are visible in at least 6 areas. Malawi, South Africa, and Bhutan are all very textually similar (< .4). Likewise, additional island groupings reappear as St. Kitts and Antigua and Barbuda are nearly identical copies according to Jaccard Distance (< .2). Barbados, Bermuda, Singapore, and Trinidad & Tobago cluster as well. Several Spanish-speaking nations, including Costa Rica and Peru, also group together. Madagascar and Mauritania form a block at the upper right-hand corner. Some recent African rules also appear relatively similar, with Kenya, Mauritius, and Botswana forming a block in the center of the heatmap. China and California are relative outliers.

Maybe most interesting is the fact that the text of the GDPR is the most dissimilar to all other laws passed in the period, indicated by the light-yellow column at the far left and the bottom-most row, even more so than Yemen. This indicates that while countries may be copying GDPR terms, principles, and ideas, they are not copying GDPR itself. Where the DPD seemed to spur some direct copying into national legislation, especially Poland and Greece, GDPR has had no such effect.

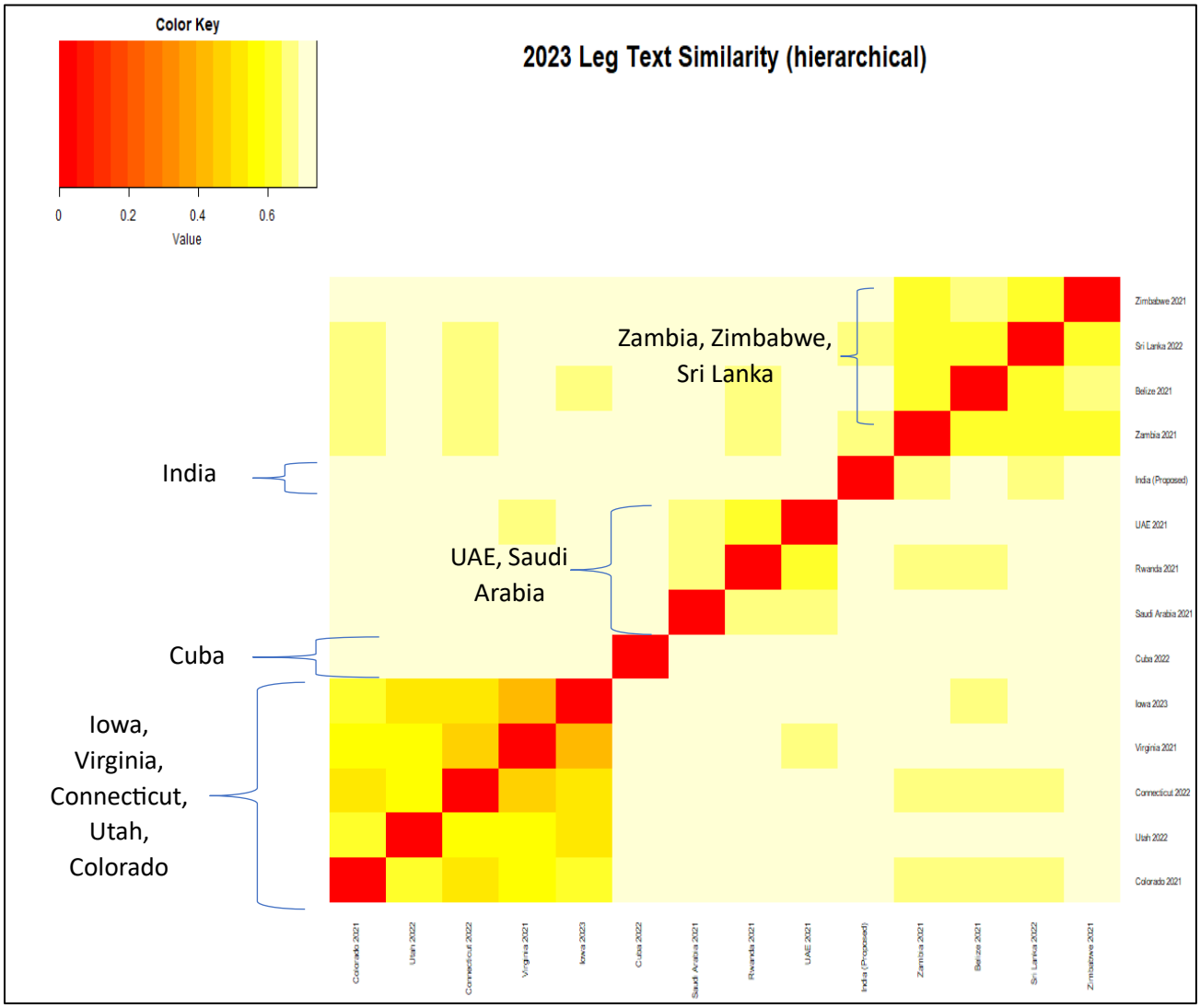


Figure 3.14. Heatmap of textual similarity measured by Jaccard Distance 2021-March 2023. Colors towards the red end of the spectrum indicate lower Jaccard Distance values, and higher degrees of similarity. Hierarchical sorting; no dendrogram activated.

In Figure 3.14 lighter yellow dyads reappear, indicating a reversal of the red shift. U.S. States are a clear grouping. At the other end of the heatmap in the upper right-hand corner, another group of East African Jurisdictions including Zambia and Zimbabwe visually stand out. Rwanda and Sri Lanka approach

this group. In the center of the map, the UAE and Saudi Arabia have similar rules. Overall, Cuba is the least similar to any recent adopters. India's proposed rule is also an outlier.

Since March 2023, existing trends have continued. Island nations (e.g. Dominica) and East African Nations (e.g. Tanzania) continue to group while American states visually stand out as a block of similar texts. Compared to Figure 3.14, the red shift has reappeared as more dyads of darker colors exist in Figure 3.15 than in Figure 3.14 on a percentage basis. But grouping, particularly around American states, seems to play a larger role than the red shift.

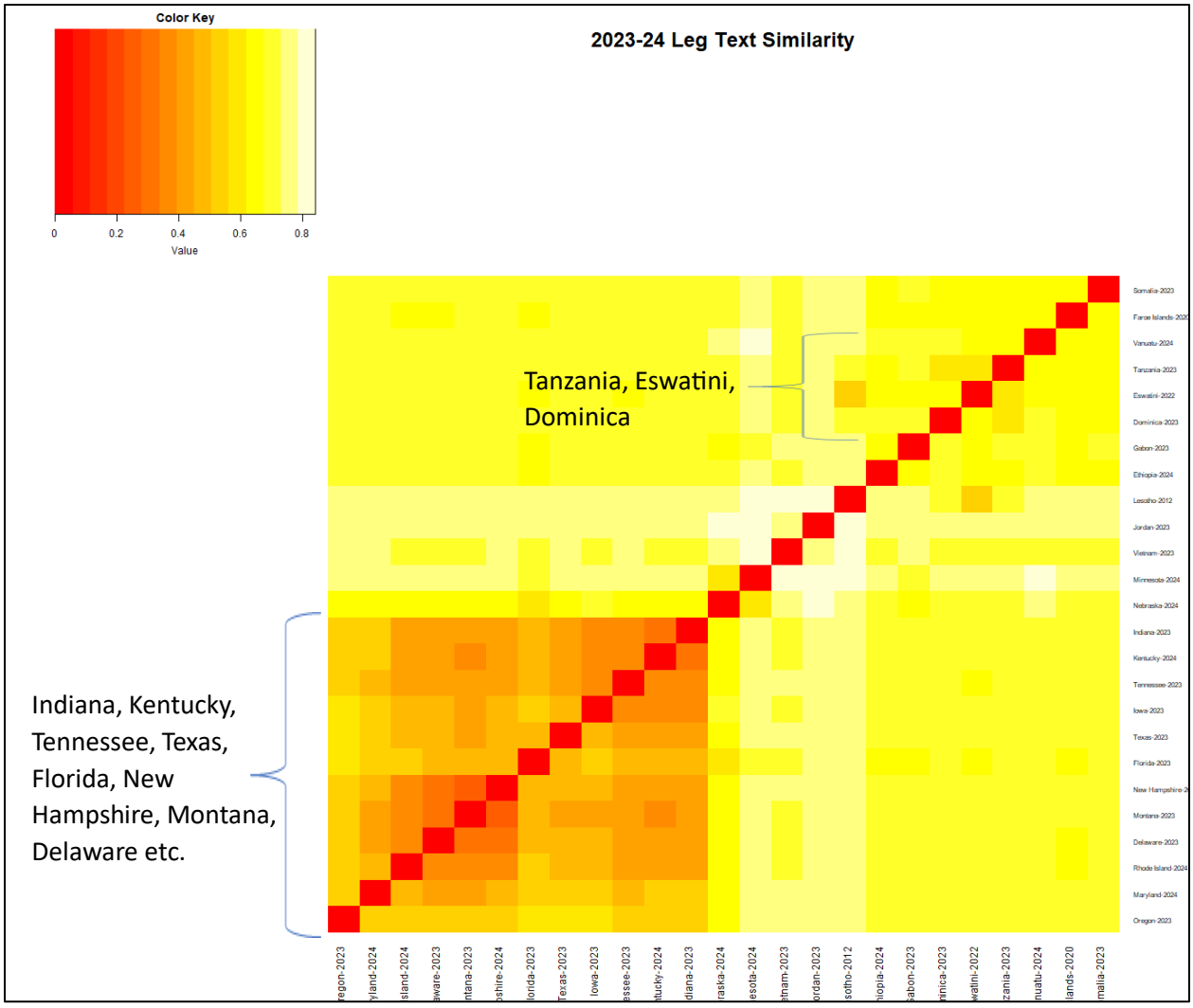


Figure 3.15. Heatmap of textual similarity measured by Jaccard Distance March 2023-September 2024. Colors towards the red end of the spectrum indicate lower Jaccard Distance values, and higher degrees of similarity. Hierarchical sorting; no dendrogram activated.

§3.5 Discussion

The heatmapping results do vindicate the convergence story. Broadly speaking, data protection laws have gotten more similar over time. In comparison to benchmarks, many dyads, particularly after

2000 show degrees of statistical distance that is below that of laws which are known copies and many more show distances at least below that of the mid-range of laws that were heavily influenced by laws abroad. Large markets seem to be immunized from foreign influence in their text while small island jurisdictions and Southern and East African nations have reused a great deal of text from their neighbors. The same is true of American states, which seem to be converging on a widely reused textual model, apart from California. But the convergence process has also taken some unexpected turns. For example, it appears GDPR has been less successful than the DPD in spurring copy-cat legislation. Instead, shared regional language, seems to have become popular. It may be that the ideas of GDPR are being copied but the language of EU legislation is not. This would help explain why the statistical grouping that the heatmaps reveal has occurred more frequently since roughly 2010 as it would point to a simplified EU-model being adopted on a regionally-specific basis. That would mirror what Frankenberg refers to as the IKEA theory of constitutions, where jurisdictions "shop" for regionally-appropriate constitutional elements from the worldwide corpus of constitutional rules. (Frankenberg, 2010).

A skeptic might argue that grouping is simply a feature of introducing more diverse states to the dataset. Doing so, on the skeptic's view, is likely to inevitably introduce a degree of diversity even if rules are undergoing a process of convergence. In other words, the grouping dynamics identified herein may simply be what convergence looks like in a dataset of diverse states rather than a new dynamic entirely. This counter-argument is seriously weakened by the robustness checks. What the German case demonstrates is that introducing jurisdictions overnight to a data protection project, jurisdictions with diverse populations, historical government styles, and economic outlooks, namely, the former states of the GDR, can still lead to convergence on existing models rather than grouping. Grouping among newly adopting states of the former GDP is conspicuously absent from Figures 3.2 and 3.3.

There is also the argument that grouping does not indicate the development of a regional model but instead independent copying from elsewhere (the EU for example) in a similar manner. That argument cannot be so readily discounted. Heatmap grouping shows shared text and comparison to benchmarks allows for the inference that the shared text is indicative of copying or at least strong influence. But who copied who and from where did copied text come from are questions the text-as-data approach cannot directly answer other than to narrow down the possible sources for copied text to other jurisdictions with whom a given law shares low degrees of statistical distance.

§3.5.1 The "Queen" Model

Nevertheless, the descriptive results outlined above do allow for the inductive generation of theory. Data protection's degree of textual similarity appears to follow a pattern visually modelled by

Figure 3.16. The x axis represents time and the y axis charts the position of the most dissimilar rules in the global corpus relative to one another. A large y axis distance between the upper-bound of the shaded region and the lower-bound of the shaded region is indicative of more textual dissimilarity than a shorter distance. Since the Figure looks like a Queen chess piece tipped on its side, the chapter refers to the model as the "Queen" model.

At the outset of the global data protection project, degrees of textual similarity are relatively low. Nations are likely aware of the legislative texts passed elsewhere but early models may be discarded as unworkable in practice. Over time, jurisdictions come to converge slowly on a set of rules that makes sense across contexts until they reach point A. At point A, whether through random chance or domestic innovation or some other factor, divergence processes reappear. Regulated entities then push back on differences as a burden on cross-border economic activity. Regulators may also begin to worry about the existence of "data havens," regulatory arbitrage, or a gradual weakening of rules. Or they may desire to reestablish the dominance of the converged-on model as an instance of perceived "best practice" for political economy reasons or to acquire prestige. These dual interests, business and regulatory, result in a harmonizing juncture, represented by the red line. This is a rule-based attempt to arrest growing divergence. The DPD is a perfect example.

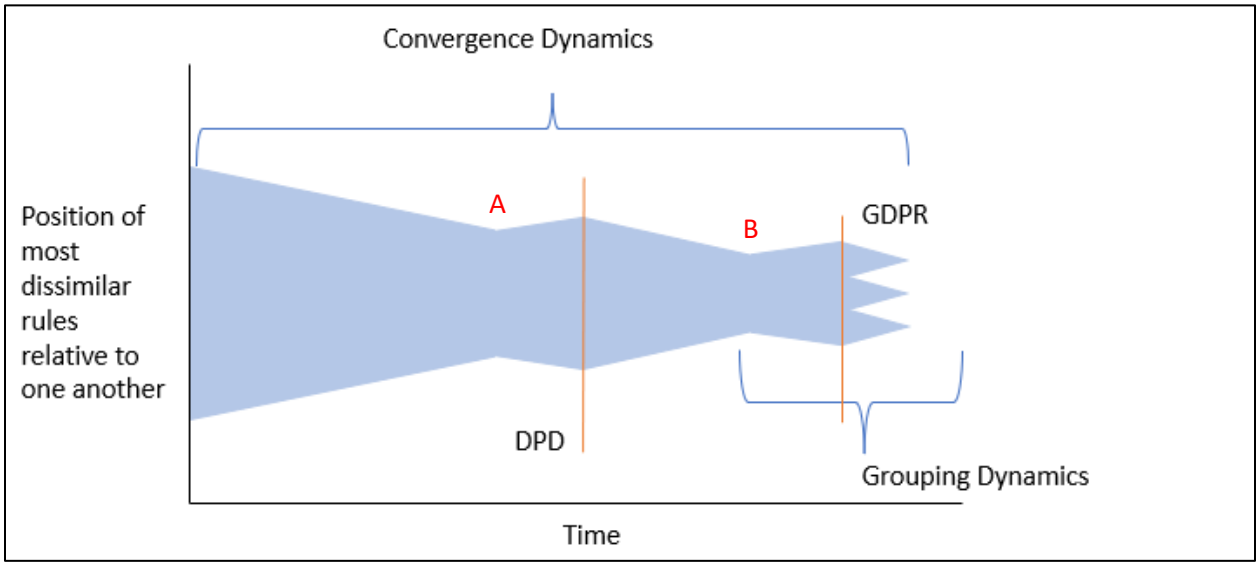


Figure 3.16. The "Queen" Visual Model of Convergence Dynamics.

After a harmonizing juncture, rules return to convergence dynamics until they reach point B, where again through chance, design, or innovation, divergent texts are introduced to the global project. As before, a harmonizing juncture occurs, this time in the form of GDPR. Only this time, possibly because GDPR is complex, technical, and lengthy, direct copying does not occur. Instead, grouping is observed in the corpus. Each prong of the Queen's crown may represent a different interpretation of GDPR parred

down or modified. It may be true that overall, the distance between the upper-bound and lower-bound continues to decline, albeit slowly, but the more germane dynamic is grouping. What causes this shift is unclear from the legislative text alone but it is visually apparent in the heatmapping.

An advantage of the Queen model is that it allows for the tailoring of research to test theory. Rather than test all rules in existence, research might, for example, focus on texts passed around harmonizing junctures. The Queen model predicts that pre and post harmonization rules have reversed-order distance values, growing before the juncture, and declining after the juncture.

In any event, the model needs further validation that cannot be provided by heatmapping alone. Text-as-data tends to need supplementation with other research methods in order to understand patterns at a deeper level than textual equivalence. (Alschner, 2020, p. 12). Heatmapping reveals the groupings, but it is crucial to validate that text is reused is legally meaningful and relevant and that copying or influence can plausibly be inferred from one legislative text to another. Contextualizing grouping may also assist in demonstrating that grouping is not due to chance use of the same language. Such issues call for qualitative, rather than quantitative analysis, which is the task of Part II.

Part II: Data-as-Text

Through three cases selected from the heatmap groupings at points critical to the Queen model, this Part address those issues. In doing so, it validates the major findings of Part I. Textual borrowing of legally relevant material (e.g., enumerated bases of processing, data subject rights, and regulatory enforcement powers) is occurring. Furthermore, textual borrowing among groups is occurring at a faster rate than, for example, from the GDPR. And outliers with large domestic markets do tend to have legally distinguishable regimes in a way that economically smaller jurisdictions do not.

Case Study 1 explores early differences between two similar European rules in Luxembourg and France. Case Study 2 explores large market outliers. Case Study 3 explores grouping dynamics prevalent since the 2010s. It evaluates the island groupings identified in Figure 3.12 and Figure 3.13 as well as the Southern Africa grouping in Figure 3.14. As such, Case Study 3 is also an exploration of what textual similarity looks like in jurisdictions with small domestic markets, providing useful contrast with Case Study 2. Within each vignette, intra-group communication links are explored to show that it is plausible copying occurred, either between the group members or possibly simultaneously from an extra-group source.

§3.6 Case Study 1: Early Divergence between France and Luxembourg

France and Luxembourg are the two most similar rules from the 1973 to 1980 subset. If they remain distinct in certain ways, that is an important validation of the Queen model since the model predicts that convergence begins from a place of relative dissimilarity in legislative texts. In 1979, when Luxembourg adopted its legislation, one year after French legislation was enacted, Luxembourg and France shared enough ties to believe that Luxembourg was at least more inspired by French legislation than other jurisdictions' legislation. To start, both France and Luxembourg share a common official language—French—and at the time both were members of the European Coal and Steel Community, Euratom, and the European Economic Community ensuring social contact between policy makers in the two countries. Both France and Luxembourg have civil law systems, a factor which comparative legal theorists often point to as facilitating diffusion of legal terms and instruments across borders (Spamann, 2009). And while German was the primary language of media in Luxembourg in 1979, most Luxembourgish legislation was and is drafted in French. Since 1984, the French version of Luxembourgish legislation is the official determinative copy for legal purposes. (Horner, 2009). Luxembourg's DPA (the *Commission nationale pour la protection des données*) also operates primarily in French and both the Luxembourgish DPA and the French DPA are members of the Global Privacy Assembly, a policy body for sharing data protection ideas, that has existed in some form since the summer of 1979. Thus, if Luxembourgish drafters in 1979 were looking abroad for best practices regarding data protection, there is reason to believe they would have looked to the French-language text of the rules passed in France just the year before.

The text of both rules reflects a degree of similarity consonant with contact, ideas sharing, and potentially copying. Both rules set up a committee, known as the Advisory Board in the Luxembourgish case, responsible for authorizing public and private data processing activities. Approval by the Committee is required prior to processing. These Committees vary in their powers and membership rules. However, the general rules regarding the collection of personal data are very similar. In Article 25, the French legislation specifies that the "collection of data by fraudulent, unfair or unlawful means shall be prohibited." The Luxembourg legislation contains the same rule in Article 14: "The collection of data by any fraudulent, unfair or unlawful means shall be prohibited." Luxembourg prohibits the collection of data concerning "political, trade union, philosophical or religious opinions" in Article 15. France has a similar rule in Article 31, prohibiting processing of "political, philosophical, religious opinions or trade union affiliations" without "express agreement," although France also includes "racial origins."

Perhaps the area where the legislative texts overlap the most is the "Right of Access," a right created by both laws that allows data subjects to access information stored about them in public and private data banks for a nominal fee. Table 3.5 illustrates a few roughly identical passages with red, italicized text indicating shared text.

France 1978 Act No. 78-17 of 6 Jan. 1978	Luxembourg 1979 Act of 31 March 1979
Article 36	**Article 22**
Any person having right of access *may demand* that data concerning him which are *inaccurate, incomplete, ambiguous or out of date or the collection, use, communication or storing of which is prohibited* be corrected, completed, clarified, updated or obliterated.	The person entitled to exercise the right of access *may demand* the rectification, completion, classification, updating or deletion of data concerning him *if they are inaccurate, incomplete, ambiguous or out of date, or if the collection, storage, disclosure or retention thereof is prohibited.*
Article 38	**Article 23**
If data have been communicated to a *third party*, the latter shall *be notified of their correction or cancellation* unless exemption Is granted by the committee.	If a datum has been transmitted to a *third party*, its *rectification or cancellation must be notified* to such third party.
Article 40	**Article 21 paragraph 3**
If the exercise of the right of access *applies to medical data these may not be communicated to the person concerned except through a doctor designated by him for that purpose.*	Where the right of access is *exercised in respect of medical information, such information may not be divulged to the person concerned except through a doctor designated by him for that purpose.*

Table 3.5. Textually similar passages in "Right of Access" rules between France and Luxembourg.

Overall, the picture that emerges is that Luxembourg appears to have selected portions of the French legislation and modified it for its own purposes. The Criminal Provisions, for instance, are roughly in line, although the maximum prison sentences are lower in the Luxembourgish case. Likewise, the French legislation begins with aspirational principles subjecting data processing to the needs of human rights, private life, and human identity. It states, in passages which presage the modern trend toward data due process, that "no administrative or private decision" should be based "solely on automatically processed data" and that "no judicial decision" should be based on "automatically processed data" either (Act No. 78-17 Art. 2). Conversely, the Luxembourgish legislation begins with functional definitions

and moves beyond natural persons to include legal persons within the scope of protection from improper data use. In a sense, the Luxembourgish legislation is the French legislation stripped of aspirational rights rhetoric and parred down to its most functional elements. Ultimately, the two do share critical passages in ways that appear unlikely to be due to random chance in phrasing or word usage. But the two legislative texts remain distinctly rooted in national legal predilections. That the two most textually similar rules of the 1970s remain relatively distinct in style and legal content confirms an aspect of the Queen Model: that the data protection convergence proceeded from a point in time with relatively high degrees of difference in text.

§3.7 Case Study 2: Large Market Outliers

Perhaps as expected, jurisdictions with large domestic markets remain outliers in their respective time brackets; California, China, and India are excellent examples. Each has a legally distinctive approach to data protection that is not shared across texts. In the case of the United States, an internal mini convergence may be sheltered by the size and power of the overall U.S. market, or as Chander et al. (2021) suggest, by the domestic leadership of California, which by itself has a higher GDP than France, the U.K., or India. Across these large jurisdictions, the story is not so much about how each shares rules with each other but how each legislative text is distinct from one another and from European-style rules. Particularly in the 2010s and 2020s, these differences may mark the inflection points A and B in the Queen Model, perhaps leading to renewed attempts to export uniform rules. Table 3.6 summarizes key differences in legislative text regarding legally important legislative elements. In comparison to the GDPR, all three large-market rules are distinct in key elements and perhaps not more so than in their diverging approach to the default legality of processing.

Element	California	China	India (proposed)	GDPR
Definition of Personal Data/Information	"information that identifies, relates to, describes, is capable of being associated with, or could reasonably be linked, directly or indirectly, with a particular consumer or household." [Ch. 55 Sec. 3]	"all kinds of information, recorded electronically or through other means, that taken alone or together with other information, is sufficient to identify a natural person's	"any data about an individual who is identifiable by or in relation to such data" [Ch. 1 Art. 13]	"any information relating to an identified or identifiable natural person ('data subject')" [Art. 4(1)]

		identity" [Art. 76(5)]		
Scope of Application	Any "business" that (A) Has gross revenue > $25 million (B) Buys, sells, or processes personal information of 50k + Cal. consumers (C) or derives 50% + of its revenue from selling Cal. personal information [Sec. 9 c(1)]	Any "construction, operation, maintenance and usage of networks" within the "mainland territory" of the People's Republic of China [Art. 2]	Processing of digital personal data within India where personal data was collected online or if offline, is digitized as well as processing outside India if connected to the "offering of goods or services" within India but in no case to non-automated processing [Ch. 1(4)]	Material Scope: Any "processing of personal data" not "outside the scope of Union law," "by Member States" regarding common foreign and security policy [Art. 2] Territorial scope: Any "controller or processor in the Union," + any processor of Union personal data who "offers" goods or services in the Union or monitors behavior of Union citizens [Art. 3]
Lawful Bases of processing	Processing is lawful	None specified	Any purpose which the "Data Principal" has consented to or can be "deemed" to have consented to [Ch. 2(5)]	Processing is unlawful unless: (1) consent acquired (2) processing necessary to contract (3) necessary to comply with legal obligation (4) necessary to protect vital interests (5) necessary to public interest or official authority

				(6) necessary for legitimate interests of controller [Art. 6]
Data Subject/Principal/ Consumer Rights	(1) Right to Disclosure (2) To Deletion (3) To opt-out of sale to third parties [Sec. 1-5]	None specified	(4) To information regarding personal data (5) To correction and erasure (6) To redress grievances (7) To nominate a representative in case of death [Ch. 3]	(1) Transparency (2) Access (3) Rectification (4) Erasure (5) Restriction (6) Data Portability (7) Objection (8) Avoidance of Automated Decision-Making [Ch. 3]
Compliance/ Regulatory Entity	California Attorney General [various]	State network information departments [Art. 8]	Data Protection Board [Ch. 5(19)]	European Data Protection Board/Independent DPAs in Member States [Art. 68]

Table 3.6. Selected provisions of outlier rules in comparison to the GDPR.

In contrast to the case of France and Luxembourg, the emergent picture here is that each jurisdiction is making its own way. Only the GDPR makes processing unlawful as a default starting point while the other three begin from the assumption that processing is lawful unless it violates some specific rule against deception or processing without informed consent (or, in the case of India, "deemed" consent). Likewise, the rights afforded the individual whose personal data is at issue vary greatly across the countries. The GDPR provides, by far, the most exhaustive list of rights. China specifies no specific rights for data subjects.

Even in basic areas where one might, *prima facie*, expect some degree of overlap either due to a paucity of methods with which to express a particular legal idea or due to borrowing, the rules in these large markets tend to be different. A perfect example is the definition of personal data. While each jurisdiction tends to adopt the basic position that personal data refers to information that identifies or could identify a natural person, each takes a different path toward elucidating that concept. California and the GDPR provide a non-exhaustive list. India and China instead stick to a rote, possibly circular formula where personal data is data that tends to be personal, i.e., identify a person. (See (Schwartz &

Solove, 2011)). The same pattern of difference is observable in scope provisions. India may have borrowed clauses referring to offering of goods and services from the EU while China and California take distinct approaches. California conditions application of its rules on the size, revenue, and business practices of businesses while China has a blanket provision that applies the law to almost any conceivable entity within the borders of mainland China. At the same time, India is itself an outlier in scoping provisions in specifically exempting non-digital, offline processing of information even at scale.

The most that can be said is that legal concepts like "personal data" have diffused across borders. In California, China, India, and the EU, specific wordings and specific legal meanings appear to have remained confined to national boundaries. As much as national boundaries may function as barriers to legal wordings, they may also shelter nascent mini convergences. U.S. States are a strong example. At least four U.S.-state level dyads are among the ten most similar rules by Jaccard and Cosine Distances: Virginia and Iowa, Colorado and Connecticut, Virginia and Connecticut, and Connecticut and Iowa (see Appendix A). At the same time, these rules diverge generally from California's rules. It may be the case that the Queen Model is playing out at the sub-national level in the U.S. with California taking the form of a pioneer rule set that is influential but not widely copied, akin to Sweden's 1973 rules. Table 3.7 summarizes two cases of legally significant and nearly identical terms; California appears as the outlier in each case. Red, italicized text indicates text shared across at least three jurisdictions.

Element	California	Colorado	Connecticut	Iowa	Virginia	Utah
Targeted Advertising	No definition	*"displaying* to a consumer an advertisement that is selected *based on personal data obtained or inferred over time* from the consumer's activities *across nonaffiliated websites, applications*, or online *services to predict consumer preferences or interests"* [Sec. 6-1-1303(25)]	*"displaying* advertisements to a consumer where the advertisement is selected *based on personal data obtained or inferred from that consumer's activities over time* and *across nonaffiliated Internet web sites or online applications to predict such consumer's preferences or interests"* [Sec. 1(28)]	*"displaying* advertisements to a consumer where the advertisement is selected *based on personal data obtained* from that consumer's activities over time and *across nonaffiliated websites or online applications to predict such consumer's preferences or interests"* [Sec. 1(28)]	*"displaying* advertisements to a consumer where the advertisement is *selected based on personal data obtained from that consumer's activities over time and across nonaffiliated websites or online applications to predict such consumer's preferences or interests"* [Sec 59.1-571]	*"displaying* an advertisement to a consumer where the advertisement is *selected based on personal data obtained from the consumer's activities over time and across nonaffiliated websites or online applications to predict the consumer's preferences or interests"* [Sec. 13-61-101(34)(a)]
Controller Duty Regarding Sensitive Data	None	*"not process* a consumer's *sensitive data without first obtaining the consumer's consent or, in the case of the processing of personal data concerning a known child*, without first obtaining consent from the child's parent or lawful guardian" [Sec. 6-1-1308(7)]	*"not process sensitive data* concerning a consumer *without obtaining the consumer's consent, or, in the case of the processing of sensitive data concerning a known child, without processing such data in accordance with COPPA"* [Sec. 6]	*"not process sensitive data* collected from a consumer for a nonexempt purpose without the consumer having been presented with clear notice and an opportunity to opt out or if *processing of sensitive data concerning a known child in accordance with COPPA"* [Sec. 4(2)]	*"not process sensitive data* concerning a consumer *without obtaining the consumer's consent, or, in the case of the processing of sensitive data concerning a known child, without processing such data in accordance with COPPA"* [Sec 59.1-574]	*"not process sensitive data without* (a) first presenting the consumer with clear notice and an opportunity to opt out of the processing; or (b) *in the case of a known child, processing the data in accordance with COPPA"* [Sec. 13-61-302(3)(a-b)]

Table 3.7. Two examples of shared language across American states with California as outlier.

As an overarching pattern, Iowa and Utah, two Republican-dominated states, appear very close to each other in terms of selected text, while the text adopted in Colorado, Virginia, and Connecticut appear as variations on a theme if not exact copies of each other. Political grouping makes some sense as national political parties at the American state level operate as networks through which policy can spread across state borders. (Gray, 1973; Linder et al., 2020).[28] California, however, does not appear to the be genesis of the specific wordings. Instead, Colorado appears to be the originator with other states updating the Colorado template in ways that simplify rules, such as the replacement of the duty of a processor to acquire the consent of a parent or guardian with the duty of the processor to comply with the federal-level Children's Online Privacy Act ("COPPA") when processing the sensitive data of a known child.[29] What the similarities across American states tends to suggest is that grouping dynamics are a critical empirical result that needs to be further explored alongside general trends towards convergence.

§3.8 Case Study 3: Groupings, Islands, and Southern Africa

Grouping is also visible outside the United States (see Figs. 3.12 through 3.15). Groupings tend to show textual similarity among island jurisdictions and in Southern Africa.[30] Dyads from these regions dominate the top ten most similar rules by Cosine and Jaccard Distance (see Appendix A). It is plausible that these small market jurisdictions in close geographic proximity are copying rules from an early adopter in the region.[31] Many of the island nations that group together are also members of the GPA, including Cape Verde, Cyprus, and Malta. Many in the Western Hemisphere are members of the Organization of American States, a group that makes non-binding policy across the Americas. These include Antigua and Barbuda, Barbados, St. Kitts and Nevis, and Trinidad and Tobago. Similar networks connect Zambia and Zimbabwe, both constituent parts of the Southern Africa grouping. Each is a member of the African Union, which itself agreed to a convention on Cyber Security and Personal Data Protection in 2014 meant to spur sharing of data protection best practices across the continent, and the Southern Africa Development Community ("SADC"), which has circulated draft data protection rules. Thus, there are international networks connecting each jurisdiction to one another.

[28] Private groups, like the Uniform Law Institute, that draft model legislation can also facilitate the adoption of textually similar provisions.

[29] Another possibility is that the states are clustering around draft Washington state legislation, which Lamont refers to as the possibility of a "pax Washingtonia." (Bracy, 2024). But since those draft rules were not included, that claim cannot be directly assessed.

[30] This chapter employs the regional characterizations of the African Union. Eastern Africa includes, for example, Kenya, Uganda, Sudan, Ethiopia, Madagascar, Mauritius, and Rwanda. Southern Africa includes, for example, Botswana, Malawi, Zambia, and Zimbabwe.

[31] Jurisdictions today enjoy unparalleled access to the rules passed elsewhere as global networks of DPAs continue to expand. (Greenleaf, 2015b). The Global Privacy Enforcement Network ("GPEN"), which arose out of an OECD recommendation in 2008, and the Global Privacy Assembly ("GPA" formerly ICDPPC) are two of the most prominent.

Textually, the Island Grouping and Southern African Grouping illustrate nearly identical rules that take on EU terms but are not direct copies from the GDPR. There is much clear EU inspiration but it would be more accurate as a matter of textual similarity to suggest that each of the countries have copied the EU in a similar, simplified way or that each is copying a pioneer within the region who was inspired by the EU. Table 3.8 illustrates some selected examples of regional versions of EU ideas. Each category of idea is integral to the EU regulatory architecture and therefore integral to the intentional replication of the EU model. For example, lawful bases for processing and the establishment of principles like lawfulness, fairness, and transparency are key to the legal architecture of GDPR. (De Terwangne, 2020). Likewise, Data Protection Impact Assessments are a key co-regulatory compliance system built into the EU regulatory scheme. (Bygrave & Yeung, 2022, pp. 141–143). Each is therefore a good test of whether EU rules have been copied. And it appears they have in a regionally-specified manner. In other words, each jurisdiction appears to have made a series of small amendments, leading to minimal variation on an otherwise borrowed idea.

EU Key Idea	Island Group Equivalents	Southern Africa Equivalents
Lawful Bases for Processing	Antigua and Barbuda: (1) Consent (2) To perform a contract (3) To comply with a legal obligation (4) To protect vital interests of the data subject (5) To administer justice (6) To exercise a function conferred on a person by law Barbados: (1) Consent (2) To perform a contract (3) To comply with a legal obligation (4) To protect vital interests of the data subject (5) To administer justice (6) To exercise a function conferred on a person or public body by law	Zambia: (1) Consent (2) To perform a contract (3) To comply with a legal obligation (4) To protect vital interests of the data subject (5) Public interest (6) Legitimate interests of the data processor Zimbabwe: (1) Consent (2) Implied Consent (3) To comply with legal obligation (4) To protect vital interests of the data subject (5) Public interest (6) Legitimate interests of controller

	(7) Legitimate interests of controller Trinidad and Tobago: none specified St. Kitts and Nevis: (1) Consent (2) To perform a contract (3) To comply with a legal obligation (4) To protect vital interests of the data subject (5) To administer justice (6) To exercise a function conferred on a person by law Malta: (1) Unambiguous consent (2) To perform a contract (3) To comply with a legal obligation (4) To protect the vital interests of the data subject (5) To carry out an activity in the public interest (6) Legitimate interests of the controller	
Data Protection Impact Assessments Required When	Barbados: (1) A new technology is likely to result in "high risk to the rights and freedoms of an individual" Trinidad and Tobago: (1) A public body project would "substantially or materially impact personal information" Other Island Jurisdictions: none	Zambia: (1) Automated processing with legal effect (2) Large scale sensitive personal data processing (3) Systematic monitoring of a public area Zimbabwe: none

Processing "Principles"	Barbados: lawfulness, fairness, transparency, accuracy Cape Verde: transparency, respect for privacy and freedoms Cyprus: fairness and lawfulness St. Kitts and Nevis: transparency and accountability	Zambia: lawfulness, fairness, transparency Zimbabwe: trust, quality, fairness, lawfulness

Table 3.8. Island and Southern African Group borrowings from the EU.

One can see how the EU idea is translated into a distilled, widely shared idiom within the group. Data protection impact assessments, which occupy almost a full page chock full of detailed rules in the GDPR are reduced to a few clauses in the Island and Southern African groupings and then only to specific instances in Barbados, Trinidad and Tobago, and Zambia. Other jurisdictions omit impact assessments entirely. Likewise, while fairness, lawfulness, and transparency principles—all central principles of the EU data protection model—are ubiquitous across rules, there is variation in the mix of principles. Zimbabwe emphasizes trust in the information environment while Barbados touts accuracy of data and Cape Verde emphasizes respect for privacy. At the same time, all laws adopt the basic position of the EU model that data processing is unlawful unless one of a list of specified lawful bases of processing applies. These show much overlap with EU rules (e.g., consent, vital interests of data subject, legitimate interests of processor that do not outweigh rights) but also show in-group similarity not present in the EU list. For example, it is common among the Island Group to specify that processing is lawful if its purpose is to "administer justice" or to "exercise a function conferred on a person by law." These formulations are absent from EU rules. So, while it appears true that these groups are converging on EU ideas and a broadly EU model, arguably the more salient dynamic is the grouping around shared ideas. These groupings confirm the empirical reality of the tines of the Queen's crown.

§3.9 Conclusion

There are several important takeaways from the empirical work in this chapter. Undeniably, the CDP convergence story outlined in Chapter 2 is supported by textual similarity patterns over time. First, one can observe the shift toward the red end of the spectrum on heat maps from 1973 through 2010. The average Jaccard and Cosine Distances also decline over time from 1973, even if only slightly. Second, there is clear evidence of text reuse and textual similarity consonant with benchmark textual distances related to rules which are known influences on each other. At the very least, many of the dyadic distances between data protection rules tend to show degrees of similarity consonant with benchmark

values of known influence. So, while convergence is occurring—the overall distances between most rules is decreasing—the degree of similarity is also high, meaning convergence is on a narrow set of formulations. Third, the number of rule dyads with distances comparable to instances of known copying are increasing over time as shown by the increasing prevalence of dark red dyads in mapping and the prevalence of recent rules in the most similar dyads (see Appendix A). These textual characteristics are the hallmarks of a rule monoculture.

However, this chapter has also uncovered new grouping phenomena. This can be observed by the reemergence of lighter yellow tones in the 2020 and 2023 time brackets alongside the darkening of certain blocks of dyads. Case Study 2 and 3 explored grouping dynamics at the sub-national and small market level, concluding that large market size may shield emerging divergent rules from pressures toward conformity. Certainly, large markets appear to be outliers. Furthermore, while the EU remains influential, it is not the case that jurisdictions are transposing GDPR wholesale into domestic legislation. Instead, as Case Study 3 and Figure 3.13 show, if countries adopt EU-style language and legal meanings they do so in a simplified "template" manner. Jurisdictions with template rules are more similar to each other than they are to the EU model as a matter of textual distance, and arguably as a matter of legal significant as well. For example, the inclusion of compliance with a legal obligation as a lawful basis for processing in Zambia and Zimbabwe has everything to do with the legal obligations imposed by the systems in each country and nothing to do with what a "legal obligation" consists of in the EU.

This chapter proposed the "Queen" Model of textual similarity over time to help explain these results. As Case Study 1 shows, rules begin from a position of relative difference. Over time, convergence dynamics mean most rules become more textually similar. But at some point, differences reemerge. These differences prompt attempts by regulated entities and risk averse regulators to reimpose uniformity and restart convergence processes. After the DPD, convergence patterns seem to have largely reappeared unchanged (see Figs. 3.10 and 3.11). Around GDPR, however, a different dynamic seems to have become prevalent, namely, grouping and in some jurisdictions outright copying of rules developed elsewhere but crucially not EU rules. Chapter 6 explores these groupings further, addressing questions like why they exist and what they look like in practice.

Regarding the CDP literature, textual similarity analysis alone cannot reveal the mechanism driving increasing similarity, only that similarity is indeed increasing. But textual similarity does provide a few clues as to the mechanism behind convergence. Not only is local idiosyncrasy not maintained outside of large markets (with some exceptions, e.g., the Island Grouping), the existence of large markets as outliers would appear to suggest that coercion through economic means is driving convergence as

large domestic markets least susceptible to economic coercion appear most textually dissimilar. Moreover, it is not the case that GDPR is "easy" to copy into domestic legislation. It is the *least* textually similar legislative text compared to contemporaneously passed rules. So, the idea that convergence is driven by copying of the GDPR because GDPR is either best practice or copying it is the easiest way to acquire an adequacy decision would tend to be empirically dubious (but see (Schwartz, 2019)). But, at base, identifying the political economy reasons for convergence requires separate research.

While the remaining chapters address such important questions, there are also avenues for future research that this dissertation does not address. An important point to note is that this chapter only considered relatively simplistic vectorization techniques. A future approach could apply, for example, a TF-IDF (Term Frequency-Inverse Document Frequency) approach which weights n-grams for the relevance of the n-gram to the document compared to other documents in the corpus. This would clearly identify instances of domestic legal innovation as repeated phrases that do not appear in other texts would stand out in computational analysis. Another way to build on this chapter would be to increase the number of benchmarks to create a better sense of the range of possible values in cases of known influence or known copying. Heatmaps could also explore more targeted provisions of a rule rather than entire rules or analysis could be based on explicitly cutting the dendrogram at different heights rather than the visual analysis of clustering. Finally, other distance measures such as the Tversky Index or Levenstein Distances might be good supplements to the Jaccard and Cosine-based approach taken here. AI-based word embeddings approach may also be a good option for future research. Work outside this dissertation will continue to explore such possibilities.

Chapter IV

Comparative Data Protection Theories: Assessing Explanations for Rule Monoculture through Realism, Idealism, and Incrementalism

One place to begin assessing the reasons behind the textual similarity dynamics uncovered in Chapter 3 is existing theories of rule interdependence in both the CDP literature and across the fields of comparative law, international relations, and political science. At a high level of generality, existing explanations across these fields form a tripartite dialectic. When applied to CDP, that tripartite dialectical frame allows three theoretical positions or "camps" to be distinguished. These camps provide plausible theoretical explanations across four dimensions of CDP: mechanism of convergence, a jurisdiction's reasons for adopting data protection rules, degree of convergence, and efficacy of the global regime (see Table 4.1). The three stylized camps—realism, idealism, and incrementalism—reflect theoretical positions an observer of data protection's diffusion could plausibly hold.

One side of the stylized tripartite debate can be considered "realist." CDP "realism" characterizes the enactment of data protection as a strategic choice by nations in an anarchic, conflictual relationship to one another. Data protection's diffusion is consequently viewed by a realist as the result of unilateral power or coercion. States not subject to direct coercion emulate data protection rules because they have little choice due to overwhelming economic pressure on domestic entities. The realist position tends to deprioritize motivations for convergence; the exercise of coercive or unilateral power renders choice an inapt concept. Jurisdictions do not select a data protection rule set, they are compelled to adopt a particular one. Since self-interest and coercion are behind the global data protection regime, realism sees convergence and spread not as evidence that data protection has "succeeded" but as evidence of the efficacy of market power.

There is also an "idealist" position. While the idealist position acknowledges the possibility that states adopt for self-interested reasons, it argues that the realist view does not provide a full accounting of the reasons behind data protection adoption. An idealist holds the minimum position that some combination of rule quality, a genuine concern for personal privacy, and/or the dynamics of dialogue and negotiation (often with the EU), is behind adoption. Idealism links these motivations for convergence to mechanisms of convergence which prioritize bilateral relations and cooperation. For the idealist, convergence is evidence of shared consensus on a single data protection template as "best practice," with attendant positive effects for data protection's dual goal of promoting trade while protecting personal privacy via rights-based empowerment.

	Mechanism of Convergence	Primary Reason for Adoption	Predicted Degree of Convergence	Efficacy of Global Data Protection Project
Realism	Domestic Interest Group Politics/Adequacy Decisions/Coercion	Economic Pressure*	Highest	Low
Idealism	Cooperation/Bilateral Negotiation	Best Practice	Second Highest	High
Incrementalism	Policy Learning	Mixed	Lowest	Mixed

Table 4.1 CDP theories and their elements. *Asterisk indicates de-prioritization

A third theoretical position exists between realism and idealism, which can be thought of as "incrementalism." Incrementalism rejects idealism's position that a high degree of convergence is indicative of the success of the global data protection project. However, it is neither necessarily pessimistic about the efficacy of data protection, as the realist position is, nor does it share realism's narrow conceptualization of mechanisms behind adoption. Incrementalism also rejects idealism's position that consensual mechanisms are largely the cause of convergence. Instead, it sees policy learning as a major factor that explains convergence on EU-style rules where unilateral power falls short as an explanatory mechanism.

The chapter fleshes out the realist position in Section 4.1, the idealist position in Section 4.2, and the incrementalist position in Section 4.3. The goal is not to suggest that the political science theories addressed in each section necessarily lead to the three positions nor is the goal to illustrate an exhaustive or mutually exclusive typology of possible theoretical positions.[32] But in laying out the three positions, the chapter highlights the dangers of relying entirely on existing theory to explain the dynamics of textual convergence and in using the dynamics of textual convergence to try to resolve the tripartite debate in favor of one theory or another. Nevertheless, realism, idealism, and incrementalism provide the theoretical foundation for further empirical research such as that undertaken in Chapters 5 and 6.

[32] One additional note: None of the authors cited for each position wholly align with any of the positions. This is partly because CDP is a constructed *sui generis* field, so no existing author falls neatly into the various quadrants an observer of the constructed field might divide it into. Consequently, scholars and commentators' words are not employed to identify them as realists, idealists, or incrementalists.

§4.1 The Realist Position

Realism, in reference to CDP, accepts the basic premises of realism as a theory of international relations. In the realist view, the world is sorted into entities with a high degree of in-group cohesion. These entities—most often states—compete with others to acquire resources, power, and security. This competition is anarchic background and only sporadically subject to rules and values since there is no referee of inter-entity conflict. (Wohlforth, 2009, pp. 133–134). Thus, CDP Realism views professed interests in privacy through the lens of what it brings to powerful states in tangible resources or prestige derived from the ability to dictate preferred policies.

Realism is unsurprisingly pessimistic about the efficacy of data protection rules. If convergence exists, it is either the result of an equilibrium outcome of state strategies to adopt or not adopt certain data protection rules or the result of the imposition of certain rules by powerful states on less powerful states. In Bennett's words, one primary CDP question is whether the diffusion of data protection rules constitutes "the diffusion of privacy *protection*, or the diffusion of privacy *law*?" (Bennett, 2018, p. 244). Realism answers resoundingly: The global march of data protection is merely the diffusion of privacy *law*. Realism looks at the world, at the serious possibility that re-identification analysis means that data can never be stripped of identifying markers (Ohm, 2010), at the stubborn outlier of the U.S.'s notice and consent model, at the growth of surveillance capitalist logics (Zuboff, 2015), and at the growth of technologies the FIPs were never designed for (Hartzog, 2017, p. 966), and sees the global data protection project as a failure. Any gradual, halting increase in standards is dwarfed by "an almost concomitant increase in the amount of personal information collected, processed, and distributed." (Bennett & Raab, 2006, p. 283). The realist believes that it is "naïve" to think that the surveillance and information capitalist ethos will retreat by virtue of data protection rules; "in fact," wrote Flaherty, "one unintended consequence" of the presence of data protection law is that "the public has a false sense of security," due in part to the fact that DPAs "have, or have used, limited power." (Flaherty, 1989, p. 11). The only value of data protection is symbolic and expressive. At best, data protection may manage, but not ultimately halt, the "voracious and inherent appetite of modern organizations for more and more increasingly refined personal information." (Bennett, 2018, p. 244).

To make these arguments realism makes several theoretical moves. First, the realist accepts that there is a link between domestic politics and international politics. In this view, the "pressures of globalization filter through domestic institutions, which then shape state preferences" leading to international outcomes that "vary according to the problem to be solved and market power." (Farrell & Newman, 2019, p. 22). Thus, the strategies chosen in the data protection arena—for example the choice

to impose a preferred formula on another state—are in part the result of the preferences, needs, and political strategies of domestic actors. As Putnam recognized, international politics might be conceived of as a two-level game in which "domestic groups pursue their interests by pressuring the government to adopt favorable policies, and politicians seek power by constructing coalitions among those groups." In turn, "at the international level, national governments seek to maximize their own ability to satisfy domestic pressures, while minimizing the adverse consequences of foreign" choices. (Putnam, 1988, p. 434). Realism sees the existence of the two-level game as a multiplication of the leverage that the anti-data protection lobby, often consisting of business interests pushing for self-regulation, can assert. If they "lose" domestically, they can always try again in an international forum and vice versa. This argument receives some support from the American corporate law context by analogy. Roe, for example, argues that the threshold determinant of American corporate law has much to do with the position of the rule chooser—Delaware—in a federal political structure. (Roe, 2005, p. 2542). A similar argument might be made about data protection rules. States (and subnational jurisdictions) know that if they pass rules that are too stringent, regulated entities can turn to federal entities, supranational entities like the EU, or international entities like the OECD or APEC, and successfully lobby for rules that are more favorable.

Realist pessimism at the domestic level of the game is grounded in public choice theory. As Olson showed, concentrated interest groups tend to do better in the democratic political process compared to dispersed interest groups that face a small harm borne by many. (Olson, 1965). In comparison to the anti-data protection lobby the domestic pro-data protection lobby in adopting states is usually weak and dispersed as the harms of privacy intrusions are widespread and often shallow. For many, the problems that arise from a lack of data protection may seem hypothetical. (Bennett, 2010; Regan, 1995). Moreover, data protection often trades off with other civil liberties issues like government transparency, meaning a new ad-hoc coalition of civil liberties groups must be created every time data protection is legislated. (Bennett, 2010). The difficulty of building a sustainable coalition across diffuse interests means that data protection laws are often a reactive response to a perceived privacy failing or a privacy scandal like Cambridge Analytica, which catalyzes public awareness and coalition formation. Data protection law is, as a result, unlikely to be proactive.

This lack of proactiveness is compounded by the fact that there is a free-rider problem at the coalition building stage. The benefits of data protection legislation are non-excludable and because of the extraterritorial nature of data protection's scope and the possible extraterritorial nature of compliance, many groups may feel that it is sufficient to let other jurisdictions legislate and/or to wait until another group incurs the costs of coalition building. In contrast, the benefits of not having a data

protection regime or of having only a weak self-regulatory regime are concentrated in industry groups who would otherwise have had to bear compliance costs. Often self-regulation is excludable. For example, an industry could set up a "privacy seal" program available only to industry in-group entities.

The anti-data protection lobby also often has an ally in governments as government bureaucracies and surveillance agencies rely on data. Flaherty wrote that "bureaucrats are … a major source of government initiatives for information collection" because of what Flaherty called the "data imperative:" the "standard delusion" that more data will solve governance problems. (Flaherty, 1989, p. 13). Flaherty's view is borne out by sociologists Fourcade and Gordon who view contemporary statecraft as an exercise of "dataist" methods including nudges, predictive analytics, and data mining with pernicious consequences for the sovereignty of the state in comparison to private organizations as well as for the autonomy and dignity of the citizen. (Fourcade & Gordon, 2020). Flaherty concurs, arguing that "new uses for automated information" will always develop once the capacity to collect and catalogue the information is within the power of the state, very much in contravention of data minimization and consent-based collection principles inherent in many data protection laws. (Flaherty, 1989, p. 375). Reidenberg refers to that dynamic as "data creep." (Reidenberg, 2000, p. 1323). Of course, opposition political forces may have reason to call attention to government overuse of personal data, but if there is a chance that they might one day be in power, they have little reason to fully dismantle government data operations. The dearth of political will for strong data protection may extend to DPAs. For one thing, many recently enacted rules "fail to create genuinely independent data protection authorities" that could push back on government data practice (Greenleaf, 2022, p. 8). For another, DPAs depend on communication flow from other parts of the state to do their job and are, furthermore, usually funded by those other parts of the state. From a public choice perspective, all these factors render the passage of strong data protection legislation a dubious prospect.

Weak data protection regimes can occur in a few further ways. For example, the realist position can point to the fact that industry may support a government agency that is traditionally more friendly as the DPA (e.g., the U.S. Department of Commerce rather than the Federal Trade Commission). (Reidenberg, 2000, p. 1363). Another method is to frame vague rules in a favorable way during the process of compliance. As Peter Swire notes, over the last 20 years "there has been a steady increase in the professionalization and institutionalization of privacy protection." (Swire, 2013, p. 851). Vatanparast argues that such professionalization "results in depoliticization, or the removal … of political concerns to ostensibly non-political, technical, and neutral domains." (Vatanparast, 2020, p. 3). The primary of technicality, in the realist account, paves the way for industry insiders to define data protection regimes,

reinforcing a neoliberal status quo. (Johns & Joyce, 2014); (Coll, 2014). Waldman, for example, argues that this process of "neoliberal managerialism" leads to compliance by symbolic structures that do not vindicate individual privacy. (Waldman, 2019, p. 19).

All that raises the question of why data protection legislation passes at all. Public choice theory would seem to predict a dearth of legislation rather than the welter of rules that presently exist all around the world. The realist position resolves that paradox by viewing legislation as primarily prompted by the desire of domestic industry and allied groups "to reap the commercial and administrative benefits that stem from information technology and data usage." (Bennett & Raab, 2006, p. 241). A data protection regime is, in the realist sense, a legal "greenlight" to collect and monetize data. And entities need not fear the rules in most cases because they are able to shape passed legislation as they see fit by emphasizing some requirements over others all the while claiming "that they are faithfully fulfilling their obligations according to a set of criteria that are defined in terms of their own interests." (Bennett & Raab, 2006, p. 251).

The realist position also has a characteristic view on trends in the stringency of passed rules. Broadly speaking, the realist position views jurisdictions as engaging in a competition for mobile capital through the nature of their data protection rules, what Lazer refers to as a "competitive mode" of rule interdependence. (Lazer, 2006, pp. 457–458).[33] While it is empirically difficult "to find clear evidence of corporate-location decisions being driven" by the presence, absence, or stringency of data protection rules (Bennett & Raab, 2006, p. 278), the realist need argue only that the threat of a race to the bottom means that the laws that are enacted are weak. Firms need not actually move to coerce a jurisdiction into passing weak data protection rules or not passing data protection rules at all. They only need to be able to plausibly threaten the relocation of investments and jobs to "pressure states into converging on liberal [data processing] policies that favor economic exchange." (Farrell & Newman, 2019, p. 20).

Of course, the idealist position responds that many organizations have purported to adopt EU rules even when they do not have to, meaning that any competition between jurisdictions moves rules towards more stringent EU rules. It may be the case that the costliness of adhering to multiple standards means de facto, voluntary export of rules, undermining the salience of the public choice dynamics realist highlights. (Bradford, 2015). Microsoft famously took out a full-age ad in the New York Times, contending

[33] A classic example of a "race to the bottom" is American corporate law. (Cary, 1974). However, the degree to which jurisdictions are racing to the bottom in the American corporate law context is likely overstated for several reasons. First, Delaware has so clearly "won" and the marginal tax revenue gain in attracting incorporators is so small that states "take at most half-hearted steps" to compete (Kahan & Kamar, 2002, p. 685). Second, even presuming that race behavior is occurring, it is unclear whether it is a race to the bottom or in fact to the top. Delaware may simply have the clearest rules as opposed to the weakest. (Fischel, 1981; Roe, 2005).

that applying EU rules across the board was good for business to the extent it engendered consumer trust. (Rustad & Koenig, 2019, p. 395). But the realist position has a response: The idea that it is difficulty or costly to apply different rules is often inaccurate. When the cost of differentiation is low, as it often is when writing more than one privacy policy for example, the empirical evidence suggests that companies do so, abrogating any race-to-the-top dynamic. (Frankenreiter, 2021). In general, it is cheaper to comply with "different privacy standards" than it is to, for instance, build cars to different process standards. (Bennett & Raab, 2006, p. 280). Moreover, the same incentive to push for the most stringent incentive a powerful, multi-jurisdictional entity faces in all jurisdictions, namely acquiring comparative compliance cost advantage over competitors who, because of their non-transnational nature, are complying with a weaker standard, is also an incentive to lobby very hard to keep the standards low across all jurisdictions, meaning a race to the top will not converge on a particularly stringent standard. (Lazer, 2006, p. 462). In short, powerful anti-data protection groups have strong incentives to weaken legislation at the domestic level whether the global dynamic is a race to the top or a race to the bottom, at least on the realist view.

Turning to the mechanism behind the spread of rules across jurisdictions, the realist position prioritizes "market power." "Market power" is conceptually coercive and penetrative (Bennett, 1992, p. 5), so the realist position sees the mechanism driving convergence as primarily force or at least unilateral power. (Farrell & Newman, 2019, p. 21). In quintessentially realist terms, Bennett argues that democracies with strong ties to the EU (whether trade ties or formal membership)—countries like the U.K., the Netherlands, Japan, and Australia—were "forced to" create legislation for data protection. (Bennett, 1997b, pp. 227–228). Otherwise, they would have been excluded from the club "within which personal data could be legitimately communicated" causing economic consequences severe enough to nullify any choice the countries had in the matter. (Bennett, 1997b, p. 228). On the realist view, while early data protection laws may have acted as "exemplars" or as "instruments of harmonization," the modern dynamic is that laws, particularly the GDPR, "have acted as a coercive force, with significant economic consequences" for businesses that rely on data flow and governments that "wish to protect their domestic industries" from the consequences of non-inclusion in the data protection club. (Bennett, 2018, p. 244). Even where bargaining occurs, state preferences are operationalized through state bargaining power (which reflects market size, internal institutions, or some combination thereof) creating outcomes that reflect the interests of powerful states. (Drezner, 2008).

One specific mechanism on the realist account is adequacy requirements. Many countries have adequacy rules that prohibit transfer to third countries that do not have equivalent protections (or some specified level of protection). (Chander & Schwartz, 2023). In the EU, these rules require a finding by the

Commission that the third country has "essentially equivalent" protections. (*Maximilian Schrems v. Data Protection Commissioner [Schrems I]*, 2015). One conceptually easy way to qualify is to copy and paste GDPR or its language. As Chander et al., put it "because an adequacy ruling greatly simplifies data transfer in comparison to the more onerous options, many countries have sought to modify their laws to obtain such a ruling." (Chander et al., 2021, p. 1765). In a rule choice environment where states feel overwhelming economic pressure to adopt a data protection rule, they are essentially forced via the adequacy requirements of the EU to adopt a mini-GDPR in their domestic jurisdiction.

Another mechanism may be psychological pressure. Failing to adopt a data protection model (such as the EU model) widely perceived to be best practice may result in a failure of legitimacy on the world stage (Graziadei, 2019, p. 460). Having certain rules, among them data protection, may also be seen as evidence of the modernity of a given jurisdiction. (Siems, 2022, pp. 576–580). This may lead domestic rule choosers to feel compelled to adopt a particular data protection regime, with attendant negative consequences for the regime's efficacy. Sometimes the desire to fit in or seem "modern" leads to purely symbolic "sham" rules. (Law & Versteeg, 2013). And rules adopted for prestige-based reasons may not be functional when the knowledge and theory needed to make the transplanted rule work are lacking at the local level. (Bedner, 2013; Berkowitz et al., 2003).

§4.2 The Idealist Position

Idealism is not quite so dour about the efficacy of data protection rules. It accepts that sometimes states have little choice but to adopt a data protection rule but argues that states are often able to do so on their own terms. The EU is not the imposing power that realists make it out to be; it cares about rights. And even if the EU does sometimes "force" its laws on other jurisdictions, it does so rarely, and the harmonizing effect leads to the diffusion of high standards of data protection that are worth it. (Bennett & Raab, 2020, p. 453).

As a matter of rule quality, an idealist believes that EU rules are the best possible political compromise that could have been anticipated, largely because of the Snowden revelations, which put political pressure on the EU to act quickly. They are certainly much stronger than they could have been if company lobbying had been as successful as it was pre-Snowden. (Laurer & Seidl, 2021). Illustratively, the GDPR contains ideas from around the world and has the largest fines and broadest enforcement powers of any existing data protection rule, all backed up by the regulatory capacity of the EU and the desirability of the single market. It contains rights that EU courts take seriously. (Lynskey, 2015). So, in the advance of the EU model, idealism sees a "general trading up of standards." (Bennett, 2018, p. 240). In other words, while the realist view sees weak rules as the inevitable result of either a race to the top

or race to the bottom, the idealist position sees convergence as evidence of a race to the top with the EU model at the finish line. Linking privacy and trade "has not caused privacy to be subservient to trade," but instead has "led many countries to establish or strengthen their data privacy law." (Chander & Schwartz, 2023, p. 43). The GDPR and its adequacy requirement simply accelerated the process. Countries are able to implement mini-GDPRs, while tailoring their rules to national histories and regulatory capacities by interpreting the black letter law in local, idiosyncratic ways. "Evidence of convergence per se should not," on the idealist account, "be interpreted as evidence of a lack of autonomy." (Bennett, 1992, p. 8).

The idealist position draws support from the "California Effect" literature, so named in reference to the idea that California often drives other American states to increase the stringency of their environmental regulations. (Vogel, 2018); (Vogel, 1997). In a "California Effect" situation, environmental regulation in a large market where jurisdictional exit is expensive leads to de facto more stringent rules in many jurisdictions when it is difficult or costly to apply different rules within the same multi-jurisdictional entity. (Vogel, 1997, pp. 248–271). Since companies are de facto complying with the large market rule, the political costs of increasing domestic stringency of regulation are low and politicians readily reap the low hanging fruit. California Effects make the most sense when standards are process-related or cumulative rather than conflicting: for example, a rule imposing a .05 parts per million limit on carbon dioxide emissions does not inherently conflict with a more stringent rule imposing a .04 parts per million limit. Both could be complied with by complying with the more stringent .04 parts per million limit. In such a situation, entities that operate in many markets may have an incentive to push for the strictest standard they face in order to impose the same costs they face on competitors. (Vogel, 1997). Data protection idealism sees that process playing out in the global diffusion of data protection law. Companies voluntarily adopt strong EU rules to keep costs down.

Jurisdictions codify those rules in part because companies do not resist rules they already implement and, more importantly, because jurisdictions cooperate with the EU or share its socio-political values. The idealist points to the case of Japan. "Japan's choice has been a qualitative one. It affirmatively chose a system similar to and compatible with EU data protection law. It did so based, in part, on a judgment regarding the merits of competing data privacy regulatory systems." (Schwartz, 2019, p. 804). The idealist also points to California: "the EU did not set up a policy shop in Sacramento … It did not [lobby] the state legislature … Yet, somehow, the ideas of EU data protection made their way to the Golden State." (Schwartz, 2019, p. 817). By analogy, idealism leans on the revisionist account of Delaware's dominant position in American corporate law, which argues that Delaware attracts

corporations because of its clear, predictable rules, advantaging of shareholders, and judicial expertise, i.e., the quality of its corporate law. (Romano, 1985); (Fischel, 1981).[34] Thus, the quality of the Delaware rules causes them to diffuse across jurisdictional boundaries. Data protection idealism sees something similar in the diffusion of EU rules and EU-style regional templates. The EU model is simply a good one or at least so widely perceived as such that it is selected over alternatives. As Schwartz puts it, "the widespread influence of EU data protection reflects a success in the marketplace of regulatory ideas." (Schwartz, 2019, p. 775). Comparable models that might otherwise have been transposed, for example the U.S. model, have simply faced too much persuasive, sustained criticism for relying on fictive notions of consumer empowerment. (Cate & Mayer-Schönberger, 2013).

To be sure, adoption is not due entirely to rule quality on the idealist view. Instead, the EU is also "an accessible legal model" compared to other possibilities like the U.S.'s sectoral approach. (Schwartz, 2019, p. 775). Transposing one omnibus law like GDPR into domestic legislation may be easier than transposing many sectoral laws. In combination, ease of transposition and rule quality lead to adoption of the EU model entirely voluntarily even in those cases without direct cooperation between the EU and the adopting jurisdiction. On the idealist view, intentional coercion is simply rare.

The idealist also points to bilateral negotiation as a mechanism driving convergence on the EU rules, adding two responses to the realist argument that cooperation is evidence of the EU's coercive use of its powerful negotiating position (see (Burri, 2023, p. 97)). First, data protection negotiation does not always explicitly occur in a trade context and, second, even when there is the possibility for coercion, the EU does not employ it. Rather than impose its preferences on others, the EU shows "striking flexibility." (Schwartz, 2019, p. 804). For instance, the political branches of the EU moved somewhat closer to the U.S. model in the Safe Harbor and Privacy Shield agreements. Privacy Shield particularly "displays concessions by both sides." (Schwartz & Peifer, 2017, p. 164). Similarly, "[b]ecause both the United States and the European Union have trading partners that do not follow their model for trade and privacy, they both provide accountability mechanisms as a private alternative to broader legal mandates. Such accountability mechanisms permit organizations to opt-in to a binding program overseen by an accountability agent." (Chander & Schwartz, 2023, p. 33). These mechanisms include APEC's CBPR rules. Safe Harbor (prior to the CJEU's Schrems I ruling) also followed an accountability model that allowed

[34] The choice of Delaware as the incorporation location for most American companies may also simply be path dependent due to the impact of network effects that centralized corporate services in Delaware early on. (Sanga, 2020).

companies to opt-in rather than have the EU's rules imposed on them. [35] Even the EU model itself is a "collaborative effort" that incorporates ideas from outside Europe and is "marked by accommodation and compromise." (Schwartz, 2013, p. 1967).

Furthermore, adequacy is not a coercive force on the idealist view. While "the adequacy requirement has given the EU an important point of leverage in negotiations … its negotiators have not exercised unilateral power. Rather, they have flexibly assessed the adequacy of different legal systems as it suits the EU's goals at the time." (Schwartz, 2019, p. 807). Adequacy, on the idealist view, merely sets up processes of acculturation and dialogue that results in agreement on the merits of the EU's approach in comparison to other existing models. This is reflected in GDPR Article 50, which commits the EU to international cooperation mechanisms. Writ large, the EU reaches "good faith" accommodations subject to independent review by the CJEU. So, to the idealist, any unilateral effect is caused only by the "complex interplay" of European institutions. (Schwartz, 2019, p. 818).

The idealist position bolsters its arguments by pointing to the ability of transnational networks to transmit the EU model without the interpolation of market power. Such networks of transnational actors and data protection institutions (e.g., the European Data Protection Board or GPEN) constantly work to improve data protection around the globe. Thus, the idealist perspective borrows from the liberal institutionalist school of thought in international relations theory, which also focuses on institutions and networks as a backdrop for state and non-state action, (Stein, 2009), and the transnational turn in international relations theory, which placed states on a power continuum along with global networks and transnational entities (Keohane & Nye, 2001).

The networked nature of the world order reduces the need to rely on coercion to explain diffusion in idealisms theoretical approach. If networks of actors cooperate, they can gradually come to share transnational values, including a commitment to EU-style data protection. Schmidt refers to this as a constitutionalizing effect of networks. In her words, growth of "cooperative networks … gradually leads to stabilization and harmonization of internal normative values and structural settings." (Schmidt, 2018, p. 72). Rather than leading to the establishment of a New World order, to use Slaughter's terminology, "[r]egulatory cooperation can develop into more cohesive network structures which create partial constitutionalization effects" without coercion. (Schmidt, 2018, p. 95). As the EU model is adopted by

[35] The European Commission and the American FTC and Department of Transportation agreed to a new version called the Data Privacy Framework in early summer 2023, though the new agreement excludes telecommunications companies and financial services. (Walter, 2023).

more and more countries, it becomes the common language of privacy around the world—a constitutionalization effect—and that common language underlies and encourages convergence.

§4.3 The "Incrementalist" Position

Incrementalism sees the diffusion of data protection law around the globe fundamentally differently than realism or idealism. Rather than viewing diffusion as a question of the presence or absence of coercion, incrementalism sees diffusion as fundamentally a question of "policy learning," a process whereby ideas spread through information channels without formal legal or economic ties. No dialogue need occur. No negotiations need to be entered. Actors in one jurisdiction merely need to be able to gather information about the policy experiments in other jurisdictions, their outcomes, and then determine their own rule choice based on that information.

Textual convergence is the result of policy learning on the incrementalist view. For example, Bennett argues in incrementalist terms that for legislative adoptions in the late 1970s "convergence followed from the mutual process of lesson drawing within an international policy community," (Bennett, 1997a, p. 100). Lazer terms such a mode of rule interdependence the "informational mode." In the informational mode, "the choices and experiences of countries produce informational externalities, providing policy lessons for other countries to use." (Lazer, 2006, p. 456). Other countries benefit from observing the policy deliberations, choices, and experiences of a jurisdiction with certain rules because information is non-rivalrous and to a degree, such deliberations, choices, and experiences are publicly observable. In Bennett's terms the policy learning perspective assumes "that policymakers are aware of policies elsewhere, that they utilize that information within domestic policy debates and conflicts, and that this utilization can help explain policy adoption." (Bennett, 1997b, p. 218).

The incrementalist position relies on the presence of networks and attractiveness of learning in conditions of uncertainty to defend these assumptions. Bennett and Raab note that policy learning and debate frequently takes place in cross-national networks like the Article 29 Working Party (now replaced by the European Data Protection Board). "The world's 'privacy commissioners,' they write, "under whatever title, have evolved into an intricate network that meets annually as well as more frequently in smaller regional, functionally specific, or less formal contexts." (Bennett & Raab, 2006, p. 144). Separately, on the incrementalist view, policy learning is a plausible diffusion mechanism because rule choice is "information intensive," (Lazer, 2005, p. 54) and rule choosers are human. They lack the ability to compare and experiment with all possible rules domestically. Likewise, "the making of privacy policy takes place in anticipation of largely unknown technological developments and changes," which

produces a strong incentive to learn from previous adopters (Bennett, 1997a, p. 104).[36] Moreover, the proliferation of the Internet and other communications technologies vastly decreases the costs of information transfer. And the cost of filtering information about rule choice in other jurisdictions decreases as international epistemic communities (e.g., conferences, professional associations, academia), cross-national entities (e.g., NGOs), international organizations (e.g., OECD), and groups (e.g., Independent Regulators Group, Article 29 Working Party) filter information and deliver it in a digestible format to foreign rule choosers. (Dolowitz, 2000); (Mayer-Schönberger & Lazer, 2002); (Lazer, 2005). Domestically, such entities take relevant information about policies adopted elsewhere and filter it "through a variety of channels to top political and administrative elites." (Bennett, 1997b, p. 226).

So, unlike the realist perspective, the incrementalist perspective does not see coercion or economic pressure as key explanatory factors. Since it relies on policy learning, a volitional mechanism, it follows the idealist view that third countries adopt EU templates largely because of the track record of the EU model. But unlike the idealist position, which views "the widespread influence of EU data protection reflects a success in the marketplace of regulatory ideas" (Schwartz, 2019, p. 775), the incrementalist position is neither optimistic nor pessimistic. It relies less on the idea that the EU model is good for political or rights-based reasons and more on the idea that convergence is the result of it being shown to "work" in practice.

Consequently, the incrementalist position is not as sanguine as the idealist position about the efficacy of the global data protection project. Learning is slow and unlike bilateral negotiation, misconceptions cannot be addressed via dialogue or enculturation. As Lazer points out, the decreased information costs and information flow associated with diffusion via policy learning can just as easily lead to "decreased investments in innovation as decision makers wait for other jurisdictions to create policy solutions, decrease the 'biodiversity' of policy solutions available in the system, and increase the spread of misinformation." (Lazer, 2005, p. 53). As Lazer adds, "faddish" information "as well as good spreads in information networks." (Lazer, 2005, p. 62). Learning may also have a pernicious path dependent aspect. This would occur for instance if the number of compared policy experiments is small at each stage in the learning process or the original policy experiments entrench models that are suboptimal. Then, over time, a policy learning process may lead to convergence on a suboptimal outcome for network effect reasons and switching costs (e.g., the QWERTY keyboard). This is especially true if the "biodiversity" in

[36] A realist might respond that it also produces a strong incentive to stay with the herd by converging on the most widely adopted model for psychological reasons unrelated to rule quality. Moreover, it is a reason to write vague rules that lead to legal endogeneity to allow future interpretations to cover new technological trends.

the system is low and the optimal model happens not to be in the pool of policy experiments at the outset of the information cascade. What is more, information may be poor quality. There is no incentive to disseminate information that one's jurisdiction has selected a failed policy. If cover ups are widespread, learning takes place but is unlikely to lead to progressively better rules over time due to poor quality inputs.[37] Of course, the existence of such downsides has much to do with how fast convergence via policy learning occurs: "A world where everyone rapidly converges to "best practice" will likely have better policy outcomes in the short run than a world where everyone experiments … and then only slowly converge to best practice. However, the latter world will have more experimentation and may be more likely to produce better policy outcomes in the long run." (Lazer, 2005, p. 61).

In short, it remains crucial to specify exactly who is learning what and how agents are learning before determining normative implications. (Bennett & Howlett, 1992). Learning could take place within an agency given discretion rather than at the level of the state's deliberative or executive bodies. If that is the case then policy learning may lead to iteratively stronger, more pro-privacy data protection rules over time. As Flaherty perspicuously pointed out, the data protection regimes that work best in practice not only have "strong leadership" and "talented staff," they also have "a capacity to learn from experience." (Flaherty, 1989, p. 22). But policy learning could also take place within deliberative or executive bodies, bodies that are non-expert and sometimes captured. Or a jurisdiction could shortcut the learning process and believe (perhaps wrongly) that other jurisdictions have engaged in policy learning and decided on a certain model, and choose that model too (i.e., jump on the bandwagon) still roughly in line with the reasoning of the incrementalist view. As convergence could be the result of any of these situations, a converged-on data protection rule will not necessarily be the best possible outcome on the incrementalist view, in contrast to the idealist view.

§4.4 The Shortcomings of the Tripartite Approach

Such is the robust theoretical discourse surrounding the diffusion of laws across borders. Unfortunately, theoretical discourse can only go so far in providing an explanation of the results presented in Chapter 3. Nor do the statistical results presented in Chapter 3 neatly resolve the tripartite dialectic. Theoretical discourse needs to be supplemented by further empirical analysis if it is to provide a suitable explanation. This is for several reasons. First, none of the three positions are falsifiable and each could be true without disproving any of the others. Consider a legislator who feels pressure to

[37] In the data protection context, at least early on, there is little evidence that information was suppressed. The Swedish model of registration, for instance, quickly proved unworkable and even though it "had an enormous and direct influence" on the development of data protection regimes in other countries, "the exact details of the Swedish model have not been widely emulated." (Flaherty, 1989, p. 94).

adopt the EU model because a domestic constituency wants to be able to process the data of EU citizens (the realist view), who also genuinely believes that the EU model protects privacy more than an alternative model like the American sectoral approach (the idealist view), and in part holds that relative view because the EU model has been shown to work in practice and was the result of many years of development in Europe (the incrementalist view). Untangling proximate causation in such a case would be fraught. The belief that the EU model is the best in practice could be caused (consciously or unconsciously) by the domestic pressure the legislator feels. The belief that the EU model is best could cause the legislator to bow to the domestic constituent pressure. Or the years of development could be the reason why the legislator believes the EU model is best. All could be true at the same time. And interview data would not necessarily untangle the mess. The legislator herself may not know exactly what the definitive reasoning was or the reasoning could have been subconscious. She may also have motivations to dissemble.

In short, there is no set criteria which if found would indicate that one theory or the other is correct. The phenomenological reality lives in the subjective understandings of rule choosers, understandings which are not accessible. Even at the mechanism level, generalizing about which tripartite view is correct would be perilous. Evidence of the mechanism prioritized by each view—whether bilateral negotiation, a channel through which evidence of policy outcomes spreads, or a domestic interest in favor of adopting an EU model—does not prove or disprove the validity of a particular CDP theory. It merely would permit an inference that the mechanism has played a role in a particular case or set of cases. That highlights the fact that one theory may be more accurate in one set of cases than another but wrong in general or wrong when applied to another set of cases. While that alone is not a reason to question the value of theorizing, it does augur in favor of additional specificity.

One might imagine that such specificity could be supplied by linking each theory to empirical expectations, the presence or absence of which would tend to prove or disprove a given theoretical position. On the realist view, for instance, speed of adoption should roughly coincide with growth in global data trade participation by domestic constituencies and the degree of similarity should grow over time in lock-step with expanding market power and legal ratcheting up of adequacy requirements in early adopting jurisdictions like Germany, France, and most importantly the EU. The incrementalist position might expect convergence on a set of rules that has been shown to work in practice. Thus, the incrementalist position does expect that particularly criticized or demonstrably bad policies will not be adopted even if they bring domestic economic benefits. In broad brushstrokes, the idealist view imagines a world where data protection rules are grounded in rights discourse and adopted for reasons

related to genuine concern for personal privacy or the harms of unfettered data processing. And because of its emphasis on bilateral cooperation and dialogue, as well as accommodation, the idealist view predicts a somewhat lower degree of convergence in text and instrument than does the realist view. Moreover, since diffusion occurs at least partially through pairwise, cooperative, bilateral negotiation, the pace of diffusion is likely to be slower than in the realist view. If the idealist view is correct, any convergences should therefore reflect a middle way rather than strictly EU ideas and rules. The more the world reflects these distinctive expectations, the more credible one account is.

Chapter 3 does present some anecdotal evidence which corresponds to these unique expectations. For illustration, the realist hypothesis that convergences should centered around the rules in large markets—i.e., those jurisdictions with "market power"—is not supported. Thus, many smaller nations seem to have adopted their own unique EU-influenced templates rather than the EU model per se. And California's model has not been adopted across American states. But such evidence illustrates the key problem: It does not prove the veracity of one model over another. For example, the idealist might plausibly argue that countries copying only a simplified GDPR template are unlikely to receive an adequacy decision, and if rule choosers in said country are aware of that fact, the adoption of EU templates is unlikely to be driven by strictly economic reasons. In any case, generalizing from such anecdotal evidence in one or several cases to the mechanisms driving convergence across all cases goes far beyond the evidence presented here.

There is yet another problem with the theoretical approach. Each causal narrative, whether realism, idealism, or incrementalism, is much better at explaining why convergence occurs than at explaining why some textual diversity persists. In essence, each predicts a world where eventually all rules will converge on a single shared set of ideas and formulations. Why some regional distinctions or local idiosyncrasies should persist is orthogonal to contemporary theorizing. This theme is addressed in Chapter 6. What appears to be missing is a focus on choice architecture, not simply on the motivations of rule choosers. Choice architecture can be thought of as the range of choices which an actor could feasibly choose along with the parameters which delimit that range. In any legislative arena, the actors responsible for writing and adopting legislation face factors outside of their control, and about which they are aware, that constrain their choices. Such factors include domestic resource limitations and the inherent costs of formulating policy—always an information-intensive process (Lazer, 2005, p. 54). They also include preexisting local legislative styles and methods of structuring legal rules with which new legislation must cohere. (Lando, 2006). Both are present in the data protection space. (Andrea & Wiewiórowski, 2022; Koops, 2014; Pisa & Nwankwo, 2021). And thus, both provide plausible reasons to

criticize the focus of existing theory on the subjective motivations of rule choosers, which has led to a situation where there is rich theory but little prospect of resolving internecine disagreement.

Chapter V

The Improbability of Policy Learning in the Data Protection Space

Despite the difficulty of resolving the tripartite dialectic, the data collected on textual similarity and the network perspectives presented in this chapter undermine the inference that policy learning, defined as lesson drawing from the comparison other jurisdictions' rules and their effects, was a leading mechanism behind data protection convergence. Instead, what the results tend to show is that the data protection convergence is characterized by a degree of imitation that is not typically present where laws diffuse through learning processes. In other words, the textual similarity results and the diffusion of data protection laws do not match existing models of the similarity of rules or their diffusion under broad conditions of policy learning. The fact that the structures through which learning might have occurred are siloed and reliant on EU-based ideas further undermines the inference that learning was a predominant mechanism of convergence. That does not mean that policy learning did not occur in some cases or that there was never any lesson drawing going on. In fact, policy learning may well have characterized data protection's pre-1990s development, especially in Western and Northern Europe, (Bennett, 1992), or, as Schwartz points out, in the case of Japan (see (Schwartz, 2019, pp. 786–792)). But it does mean that policy learning should not be considered a strong candidate for explaining contemporary or recent convergence compared to the realist or even idealist models.

Part I below argues that the nature of the diffusion curve itself as well as the degree of convergence in data protection overall are indicative of larger amounts of imitation than legal texts which are known to have diffused through learning processes in the same period. It begins in Section 5.1 by conceptualizing learning as a concept distinct from imitation before outlining, in Section 5.2, what learning (or the lack thereof) would look like in observable variables of diffusion, before comparing data protection's diffusion to existing models of learning in Section 5.3. Part II argues that the knowledge and social structures through which learning could have occurred are not conducive to learning, as defined herein, adding support to the arguments developed in Part I.

Part I: Learning and the Diffusion of Data Protection Law

Learning resists easy definition because it encompasses a wide range of possible behaviors. Who is learning what from whom (or what), how, when, and with what consequent effect on the relationship of laws across borders are all questions that need to be answered. (Bennett & Howlett, 1992; Dolowitz,

2000). Political science tackles these questions through the lens of "policy learning." (C. A. Dunlop & Radaelli, 2020). The literature on policy learning thus serves as a point of departure for conceptualizing learning in the context of the global data protection project.

§5.1 Conceptualizing Learning: What it is and is Not

Most political scientists define policy learning as the "updating" of "knowledge and beliefs" about the nature of public policy. (C. A. Dunlop & Radaelli, 2013, 2020, p. 4). They agree that such "updating" is facilitated by the decreasing information costs brought on by modern technologies. Information and social contacts multiply in a globalized world. As Evans puts it, learning is "largely" a "function of the world of public policy becoming smaller." (Evans, 2010, p. 1). Or, as Neumayer and Plumper put it, the potential for learning increases with the "density of interactions" among the sources and recipients of information. (Neumayer & Plumper, 2012, p. 824).

When updating accounts for the experiences of other jurisdictions, "learning can be understood as a mechanism of diffusion, that is, an explanation of why and how policy choices in one country … are influenced by prior decisions in other countries." (Gilardi, 2010, p. 651); (Dolowitz, 2000). There are several typologies of learning that are helpful in the policy learning literature. May's classic division of learning into instrumental, social, and political learning is a good example. (May, 1992). Instrumental learning involves evidence about the performance of a given policy intervention in a specific context. Social learning is the invention of a novel way to handle a policy problem by comparison to existing solutions. Political learning regards the domestic political feasibility of a particular policy solution. Instrumental learning corresponds with what Rose (1991) and Bennett and Howlett (1992) call lesson drawing, or the process of drawing information about the outcome a policy choice from one setting to meet objectives in another policy setting.

Those involved in the process of policy learning are typically referred to as "policy actors," i.e., the range of individuals who possess the power "to influence government decisions in relation to a topical issue over time." (Moyson et al., 2017, p. 162). Information flows from outcome to policy maker either directly through observation or through intermediaries including policy networks, epistemic communities, policy brokers, and advocacy coalitions. (Moyson et al., 2017, p. 164). One categorization of policy actors in the diffusion process breaks them into internal actors, those who are considering domestic policy choices and interpreting information about them generated elsewhere, external actors, those who are considering implementing or have implemented a policy elsewhere, and "go-betweens," or those who act as information bridges between these two groups. (Graham et al., 2013, p. 684). Ultimately, policy actors turn to learning as a method of solving social problems when they face

conditions of uncertainty. (Heclo, 1974, p. 305). Under conditions of uncertainty, learning can illuminate solutions to a given social problem.

Borrowing from the traditions of the policy learning literature, learning is conceptually defined herein as a phenomenon where legislative actors, individuals with the power to affect the substance of domestic data protection legislation, take outcomes from other jurisdictions' legislation, including its cost, economic effect, political effect, and effect on personal privacy, into account when designing a domestic data protection regime to the extent those outcomes are observable. To treat learning as a mechanism for textual convergence is to suggest that such a phenomenon leads legislation choosers to update their preferences towards the EU model of data protection. The information on outcomes necessary to such updating may flow either through an intermediary, such as an expert or policy entrepreneur or foreign regulator, or it may be directly observed from publicly available information.

But learning is not always motivated by a desire to implement the best possible policy or to discover a policy that is objectively preferable. Policy makers have a range of motivations. For example, policy makers "may give more attention to examples that are geographically closer [or] culturally more similar." (Gilardi, 2010, p. 661). Furthermore, the nature of adversarial politics means that learning is often used as a screen or as a method with which to "justify organizational positions and interests." (Sabatier, 1987, p. 650). And policy actors look to political outcomes, such as policy change's effect on electoral outcomes, as well. (Gilardi, 2010). Such political outcomes may be unrelated to the quality of the policy choice from a technocratic perspective.

Nor is it always possible for well-intentioned policy actors to discern the objectively best policy choice. Sometimes the outcome of a policy choice is hard to measure or difficult to observe. Policy actors also suffer from a variety of cognitive biases. Policy actors are not always able to appropriately parse the information they acquire. (Moyson et al., 2017, p. 165). For example, the weight policy actors afford information gathered from other jurisdictions is affected by how well the information corresponds with previously existing beliefs. (Gilardi, 2010, p. 651). It is unsurprising therefore that policy learning does not necessarily lead to policy change. (C. A. Dunlop & Radaelli, 2020, p. 4). Even if policies do change, "policy learning" is likely to only be "one of many factors" that contribute. (Moyson et al., 2017, p. 165). Furthermore, policy learning does not necessarily lead to policy convergence. It can "create divergence" if, for instance, jurisdictions "learn from negative examples," policies that are suboptimal, due either to cognitive biases, information asymmetries, or fads, for instance. (Neumayer & Plumper, 2012, p. 824). So, it is important not to overstate the normative value of learning.

Nevertheless, learning does have important normative implications in comparison to alternative mechanisms of diffusion. At the very least it implies that policy actors are consciously evaluating the outcomes of policy choices made elsewhere as part of their investigation of what policies would be worthwhile domestically. Learning as a mechanism for diffusion thus stands in contradistinction to imitation, the mere copying of legislation without the deliberate and conscious consideration of policy outcomes elsewhere. (Shipan & Volden, 2008, p. 842). These two mechanisms of diffusion are mutually exclusive. The comparison of outcomes from other jurisdictions need not occur in a situation where imitation is present. Instead, the goal is simply to copy as much as is legally and socio-politically possible. Thus, where imitation is present, it can be assumed learning is not. Looking outside the context of policy learning helps to grasp the intuitive validity of this assumption: Which student has engaged in learning, one who copies an essay from a class example, or one whose essay has all the same elements as the class example, but does not share any text with the class example?

§5.2 Operationalizing Learning

What learning looks like in practice is difficult to specify in advance in a way that accounts for all theoretical cases. Neither convergence nor divergence is necessarily evidence of learning. Furthermore, neither the prevalence of objectively good policies nor of bad policies is evidence of learning. In part, this is because observable behavior can be explained by mutually exclusive cognitive pathways equally as well. The problem is not necessarily the inference that learning is occurring, it is the difficulty of "exclud[ing] [] alternative mechanisms," of policy change. (C. A. Dunlop & Radaelli, 2020, p. 14). For instance, if one were to observe a group of people suddenly opening their umbrellas, it would be wrong to assume that the reason the marginal member of the group opened their umbrella was because they felt the rain or predict rain (learning from experience) rather than because they witnessed everyone else open their umbrella (imitation of behavior). Another reason for the difficulty of specifying what learning looks like in advance is the difficulty of determining what prior policy beliefs were and thus whether they have been "updated." A major problem in the policy learning literature is thus how to determine whether a given actor or organization has learned. That is why, in Graham et al.'s words, most studies of learning use proxies rather than "direct measure[s] of learning itself." (Graham et al., 2013, p. 694).

One way to infer that an appreciable amount of learning is not occurring, however, is to measure imitation, which is easier to measure than learning. If imitation excludes the possibility of learning, and imitation is observed in practice, then learning is unlikely to be occurring. As Gilardi puts it, whether learning is occurring is "essentially unobservable;" the best that can be achieved is the detection of "systematic patterns in potential imitation," which support an inference that learning is not occurring.

(Gilardi, 2010, p. 654). One approach to modelling imitation is to estimate the likelihood of adoption of a certain policy based on the number of surrounding or similar jurisdictions that adopt that policy. (Boehmke & Witmer, 2004, p. 43). Another option is to estimate the degree to which a policy has its intended effect, for instance, controlling rising healthcare costs, on the chances that surrounding jurisdictions facing that policy problem adopt that policy. (Gilardi et al., 2009, p. 12). The more direct (and potentially novel) approach is to simply measure the similarity between policies as well as the time between their adoptions. If policies are statistically similar and the second policy was adopted shortly after the first, that comes close to measuring imitation itself. When "quick" "policy change" is observed, it is simply unlikely to "result from policy learning." (Moyson et al., 2017, p. 174). Too short of a period between adoptions, however, would tend to suggest the jurisdictions independently came to the same solution, while too long a period would tend to suggest that the later jurisdiction waited to observe the outcomes of the policy selected in the first jurisdiction. And, of course, statistical dissimilarity would suggest a lack of imitation. In essence, statistical similarity and a short but not too short period of time between adoptions can be thought of as necessary and sufficient conditions for imitation.

Certain observational expectations can be generated as a result. These exist along a spectrum from diffusion characterized completely by imitation, or pure imitation, to diffusion characterized completely by learning, or pure learning. At the pure imitation end of the spectrum, diffusion occurs with mere months between policy adoptions and the degree of similarity between policies selected is high, (Shipan & Volden, 2008, p. 843). Moreover, there tend to be more radical shifts in policy choice as jurisdictions strive to keep up whatever the on-trend policy choice is. In contrast, at the pure learning end of the spectrum, diffusion occurs more slowly, with periods of years between policy adoptions and the degree of similarity is variable. Likewise, since policy learning is "contextually dependent" and "most likely to contribute to changes in the secondary beliefs of policy actors" it is likely to contribute to incremental shifts in policy rather than wholesale paradigm shifts. (Moyson et al., 2017, p. 163). These hypotheses can be assessed by comparing the diffusion of data protection to known cases of diffusion-by-learning.

§5.3 Comparing Data Protection Diffusion to Learning

The paths of American state adoptions of a policy of gaming contracts with indigenous recognized tribes and OECD countries adopting patient-classification reforms in their healthcare systems serve as appropriate comparators for several reasons. First, they occurred during roughly the same time window many countries adopted data protection rules, roughly from 1980 to 2010. Second, the inference that learning occurred in each situation—that is neighboring states and fellow member

countries updated their policy choices based on observed outcomes elsewhere—is extensively documented in the literature. Third, each are reforms that, like most data protection rules, require legislation or at least the active input of legislatures. Each paints a picture of diffusion that is different in relevant respects from the diffusion of data protection rules. In particular, both diffusion curves evince more periods of time in which information could have been aggregated and outcomes assessed than the diffusion of data protection.

State-tribal gaming compacts originate with the 1988 Indian Gaming Regulatory Act, which requires that American states negotiate with federally-recognized semi-sovereign tribes in "good faith" over licensing and regulation of so-called Class III gaming, which includes most casino-style games of chance. (Akee et al., 2015, p. 192). As of September 2024, American states and federally recognized tribes have signed 1169 agreements permitting Class III gaming on indigenous lands held in trust. Such compacts set forth the types of games permitted on facilities located on indigenous land and for what specific period of time such games are permitted. (Boehmke & Witmer, 2004, p. 41). Gaming compacts are often lucrative for tribes and for states, who are able to secure a portion of revenue in exchange for negotiating or renegotiating compacts. (Boehmke & Witmer, 2004, p. 41). And because the compacts themselves are in the public record and the revenues from Class III gaming are observable, Boehmke and Witmer persuasively argue that states and tribes in other jurisdictions learn from previous compacts, at least where they do not have a preexisting pool of intra-state compacts to draw from for experience. (Boehmke & Witmer, 2004, p. 46). This makes some sense as states and tribes are likely to look to apply the experiences of their peers to their own situations. At the same time, since not all states have tribal lands within their borders, the sum of other experiences to learn from is constrained, keeping information costs down. Such conditions are theoretically ideal for learning.

Similar conditions exist regarding hospital financing reforms in OECD countries. In the 1980s health care costs rose much faster than GDP for many OECD countries. (Gilardi et al., 2009, p. 551). That problem created a need for major reforms. One reform undertaken was to change from a retrospective payment system in a hospital context, where providers were paid or reimbursed after the delivery of services, to a prospective system that pays providers a set fee in advance based on the diagnosis category of each patient in their care, creating an incentive to eliminate unnecessary care. "Patient classification systems," as such prospective systems are sometimes termed, also create easily comparable sets of patient outcomes and cost per patient data, but at the same time must be "adapted to culturally specific" medical treatments and systems. (Gilardi et al., 2009, p. 553). Gilardi et al. demonstrate that when the implementation of a patient classification system in another OECD country

led to “a milder rise or a decrease in public expenditures on health and when the empirical evidence supporting [that] conclusion is more reliable,” other OECD states are significantly more likely to adopt the policy, indicating that OECD states learned from each other’s experiences. (Gilardi et al., 2009, p. 568).

The cumulative diffusion curves of American states adopting gaming compacts for the first time and OECD countries adopting a patient-classification system between 1980 and 2005 (corresponding to Gilardi et al.’s data set), are graphed in Figure 5.1. The implementation of gaming compacts follows an S-curve-type diffusion pattern (see (Boushey, 2012, p. 133)). In contrast, the adoption of patient classification systems in OECD countries follows a stepwise pattern, without the rapid adoption of the innovation. Together what the diffusion curves suggest is that there is no trademark shape of diffusion, no specific kind of cumulative diffusion curve, that is indicative of learning.

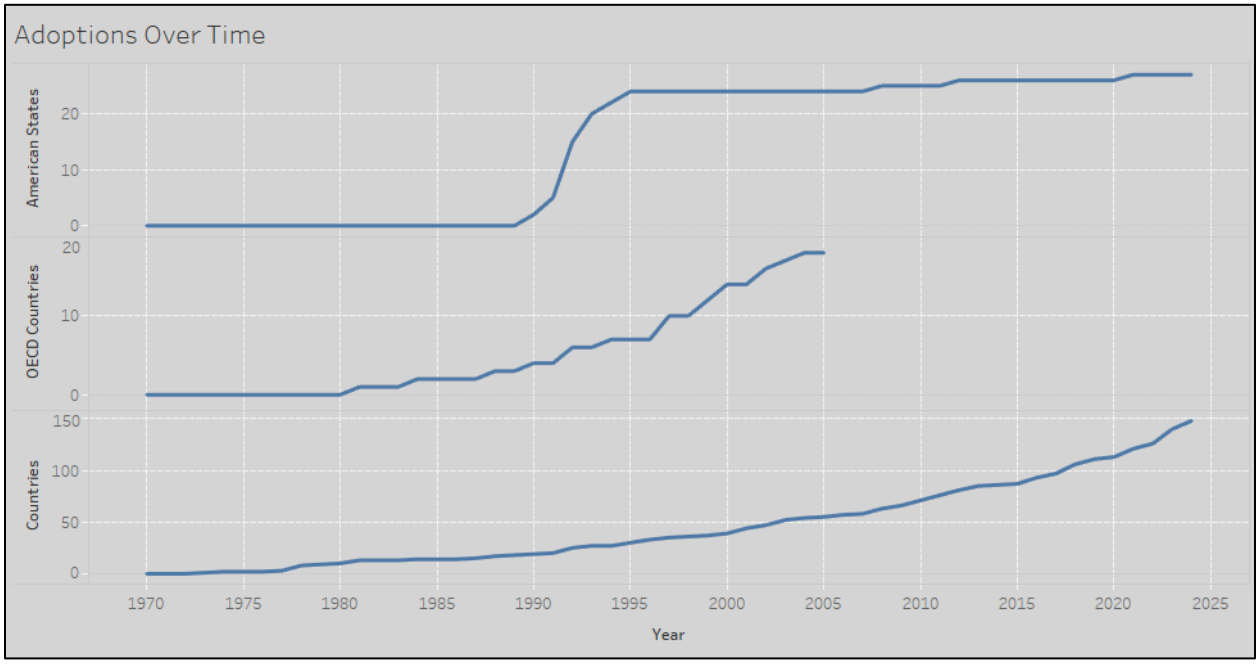


Figure 5.1. Adoptions over Time. Top panel = # of American states that have signed gaming compacts with federally recognized tribes. Middle panel = # of OECD Countries with a Patient-Classification system per year. Bottom panel = Countries adopting data protection per year.

And yet, in each case the amount of time in which the cumulative diffusion curve does not grow both on a percentage basis and as a factor of the number of years in which the cumulative diffusion curve does not grow is relatively high in comparison to the cumulative diffusion curve for data protection rules. Table 5.1 sets out these characteristics. Across the total time in which rules diffused across boundaries, data protection’s diffusion curve was only static—that is did not grow—13.5% of the time while patient classification systems and gaming compacts curves spent a considerably larger percentage

of time in periods of stasis. This result holds even when weighted for the total years of possible adoption, which helps normalize the amount of time and periods of time spent in stasis.

	# of Stasis Periods (Weighted by Total Years of Possible Adoption)	Time in Stasis Period (%)
Gaming Compacts (N = 1169)	4 (.108)	75.7%
Patient-Classification Systems (N = 19)	9 (.360)	52.0%
Data Protection (N = 148)	4 (.077)	13.5%

Table 5.1. Stasis Periods

As a theoretical matter, those periods of stasis permit the consolidation of ideas within the regulatory, academic, and rule chooser communities. They also permit the observation of the outcomes of early adopting rules. Without periods of stasis in the rules, the inference that there was sufficient time to observe outcomes, which is definitionally essential to policy learning, is significantly weakened.

The central role of time in facilitating the conditions for learning produces a corollary insight: namely, that the degree of similarity is only one symptom of the diffusion of legal rules via imitation. It is the combination of a lack of time in which to observe outcomes (see Fig. 5.1) and the existence of high degrees of textual similarity that indicate imitation. One can see these two indicators—high statistical similarity and low time—in Table 5.2. Compared to the gaming compacts case and the case of bills of rights in commonwealth countries, data protection rules have the highest percentage of dyads below benchmarks indicative of copying, as well as the lowest average Jaccard and Cosine distances across all dyads. And when dyads are similar, the amount of time between the first and second adopted rule in the dyad is much smaller than in the bills of rights or gaming compacts cases.

Explanations for these differences other than the existence of larger degrees of imitation in the data protection case than the bills of rights or gaming compact cases are implausible. To begin, these distinctions could be due to greater difficulty of producing political agreement either in legislatures or between legislatures and recognized tribes in bills of rights or gaming compacts than a data protection rule. This would naturally produce longer periods of time between adoptions by slowing down the

process of adoption. It would not, however, intuitively produce high degrees of similarity. There could be barriers to the spread of information in the gaming and bills of rights contexts but not in the data protection context but that possibility is somewhat undermined by the fact they all diffused during the same time and thus information could have theoretically traveled along the same information technology architecture. One could also imagine that it takes longer for the effect of some rule choice to become clear or occur. But there is no immediately intuitive theoretical reason why the effect of data protection rules would be so much quicker to occur and be observable than gaming compacts or bills of rights.

	Avg. Distance (Cosine) (Jaccard)	Similar Dyads (%) (Cosine < .2) (Jaccard < .69)	Avg. Years Between Top 4 Most Similar Dyads (Cosine) (Jaccard)
Gaming Compacts (n = 20)	.422 .824	7.4% 12.6%	8.75 8.5
Bills of Rights – Shared Legal Origin (n = 10)	.604 .840	0% 0%	28 38.5
Data Protection	.386 .752	1.39% 14.06%	3.5 3.5

Table 5.2. Warning Indicators of Imitation

What would tend to produce high similarity and low periods of time between the highest similarity dyads is the independent discovery of an obvious answer to a regulatory or legal problem in a world of limited rule options. This “perfect storm” phenomenon would produce the same rules in multiple jurisdictions presuming those multiple jurisdictions face the same problem at the same time. In other words, the catalyst problem the rules seek to solve confronts rule choosers in multiple jurisdictions at the same time. A good example might be the implementation of seatbelts rules. The evidence that the lack of seatbelts created problems for personal safety while driving was unearthed in several jurisdictions at once, all of whom faced a fairly obvious universe of regulatory options: mandate seatbelts or permit personal choice as to seatbelt use for broadly libertarian reasons. (Leichter, 1986). The proliferation of similar rules mandating seatbelts is therefore unsurprising and not necessarily evidence of imitation. Evidence for that perfect storm is absent in the data protection case.

Research deepening these conclusions is required. In particular, qualitative interview research with rule choosers in a variety of contexts would be useful in establishing the validity of inferences from the speed, stasis periods, and degree of dyadic similarity in a rule diffusion context. The generalizability of these findings across a broader range of examples is also a possibility for future research. What would a generalizable theory of learning versus imitation look like within the field of comparative law? That question remains to be answered.

Finally, one point of caution is necessary. Imitation as mechanism of diffusion may seem pejorative. It could unnecessarily imply that rule choosers in a particular jurisdiction are unsophisticated. That inference would be wrong. Learning is not inherently efficient and imitation is not inherently suboptimal. Under conditions of resource scarcity, imitation may in fact be the best use of regulatory resources. It is certainly cheaper than learning. And in a case where the imitated rule is optimal, imitation is significantly justified. Moreover, even half-hearted imitation can produce positive domestic changes. It may be wrong to criticize a jurisdiction for copying a human rights regime it has no intention to enforce from the UN Charter if domestic groups are able to use the implementation of that regime to call out a government for hypocrisy. At heart, whether imitation is normatively good depends on factors other than its inherent characteristics.

Part II: Learning and the Structures of Data Protection Diffusion

The nature of the networks of knowledge and social relationships through which outcomes might have spread further undermines an inference of policy learning. Focusing on the types of epistemic communities that Bennett, Newman, and other CDP scholars point to as information conduits for data protection ideas, the following demonstrates that the network structures in such communities lack small world structure, a kind of information architecture conducive to information diffusion, are characterized by siloing, the prevalence of single degree “dead-end” nodes, and, finally, that those nodes which have high centrality in the network structures are people or papers which tend to espouse the supremacy of the EU model rather than compare models. Each point suggests that the network architectures the incrementalist position might point to as evidence of policy learning in fact undermine its existence.

The following is broken into three sections. Section 5.4 introduces the network concepts employed as well as the surprising gap in the CDP literature related to network structure, before Section

5.5 addresses methods and data collection. Finally, Section 5.6 analyzes results, demonstrating the absence of "small-worldness," siloing, and the relative network hegemony of pro-EU nodes.

§5.4 The Absence of Network-Structure Analysis in CDP

Legal diffusion, defined as the spread of rules and the ideas behind them, is often a question of transnational relationships through which information is exchanged. (Josua, 2021; Shur-Ofry et al., 2019; Spamann, 2009; Twining, 2004). Traditional legal diffusion modeling borrows much from studies of the diffusion of innovations. (Engel & Weinshall, 2022); (Shur-Ofry et al., 2019). As a result, the complex dynamics surrounding the diffusion of legal ideas and the rules they catalyze are distilled down to what amounts to a counting and graphing exercise: How quickly do jurisdictions adopt a given rule and what insights does the adoption curve lend itself to. (Shur-Ofry et al., 2019). While such research generates useful insights, (Goldbach, 2019), and as Part I shows, also tends to undermine an inference of policy learning, it leaves out analysis on the nature of the channels within which information about the outcomes of rules spreads. In particular, such research misses out on the richness of network perspectives, which have provided novel empirical insights in other fields concerned with diffusion phenomena, (e.g., (Chané & Sharma, 2017; Hafner-Burton & Montgomery, 2011; Tam Cho & Fowler, 2010).

That is unfortunate given the tendency within CDP of ascribing causal power to information networks. (Bennett, 2010; Newman, 2008). For example, Newman argued that the 1995 DPD and subsequent data protection diffusion from Europe to the world was the result of strong, concentrated networks of data protection regulators within Europe with content expertise sufficient to overcome opposition within the political processes of the EU. (Newman, 2008). Fundamentally, much of the information exchange that is crucial to the idealist and incrementalist perspectives is thought to occur through subnational channels of communication between content-area experts: regulators, legislative staff, academics, practitioners, and the like. (Bennett & Raab, 2006, p. 144). Some CDP scholars argue such "epistemic communities," may facilitate convergence through the diffusion of best practices and the comparison of common experiences. (Dolowitz, 2000; Lazer, 2005; Mayer-Schönberger & Lazer, 2002). And others argue such networks may facilitate "stabilization and harmonization" of normative values within the data protection space, what Schmidt refers to as a "constitutionalization effect," (2018, p. 95), in line with the idealist picture of CDP.

But the topology of the network of data protection actors remains unspecified. In Lazer's words, the "actual *architecture*" of information diffusion remains to be specified. (Lazer, 2005, p. 64). Bennett, for example, notes that the success of the network of privacy advocates is in part determined by the

"nature and strength of the network itself," but concludes only that the global privacy advocate network is "nonhierarchical" and "composed of multiple groups," tied together by conferences, campaigns, and net activism. (Bennett, 2008, p. 195). He adds that certain "traveling evangelists," tie the "segmentary and polycentric" groups together in ad hoc policy coalitions and spread information between groups. (2008, pp. 196–197). These are fundamentally conclusions about the topology of the network but they lack empirical, rather than qualitative, description. This Part fills that gap by exploring the diffusion of data protection rules as a matter of complex and contingent network relationships in a prominent epistemic community within the data protection law and regulatory space as well as the structure of knowledge within the CDP literature.

While foreign to CDP, the study of network architecture is not foreign to socio-legal studies. Historically, it has tended to fall into three camps: social networks with the field of law (e.g., the relationships between members of the bar, (Heinz & Laumann, 1982)), the structure of law itself (e.g., cross-references in the French Legal Code, (Boulet et al., 2011)), or the structure of legal authority (e.g., the citation practices of the American Supreme Court (Fowler et al., 2007)). (Whalen, 2016). For example, De Visser and Claes analyze the transnational networks of jurists that "Europeanize" national court systems through the sharing and contesting of legal ideas. (De Visser & Claes, 2013). Dallara and Piana expand on that work by focusing on judicial networks as a matter of governance in constitutional democracies. (Dallara & Piana, 2016). As a further example, Siems examines judicial citations across borders within Europe from a network perspective, concluding that a shared language and other such nonlegal factors are a powerful explanation of citation trends. (Siems, 2023).

This type of work—using network science approaches to generate insights about law—is not confined to European law. Analyzing trends in citation of out-of-state supreme courts, Peter Harris was able to demonstrate patterns in American state court prestige and authority using a network perspective. (Harris, 1982). More recently, the spread of ideas in human rights law has been approached from a network perspective. (Dothan, 2018). Such work helps explain the internationalization of human rights discourse and the extant effectiveness of human rights jurisprudence that lacks direct enforcement mechanisms. Similarly, social network analysis has yielded insights in the study of international arbitration, a field which relies on a core group of professionals sharing ideas and best practices, (Puig, 2018), and in the study of law's role in colonial worldbuilding, (De Almeida Costa, 2023). Other research problematizes networked transnational idea sharing in the legal and regulatory space as creating the potential for the erosion (or perception of erosion) of sovereignty (e.g., (Slaughter, 2017). CDP can benefit from such approaches in several ways. It might illuminate patterns in cross-border legal

citations or in which data protection regulators follow each other on social media. Such data would help to explain why certain ideas flow to certain jurisdictions through the social and professional relationships between rule choosers.

One further area in which all legal analysis, including CDP, stands to gain much is the advances made in the analysis of real-world "small world" networks. Small world networks refer to a special class of networks with certain properties compared to a random graph. (Watts, 2004, p. 244).[38] In particular, small world networks are characterized by high clustering and short average path lengths compared to random graphs of the same basic parameters. (Telesford et al., 2011, p. 368). At the local level, such "small world" networks are characterized by triadic closure which is bridged at the global level by "weak" ties that serve as connecters between otherwise unconnected components of the network. (Watts, 2004, p. 246); (Milo et al., 2002). In small world networks the degree distribution, a histogram of how many other nodes each node is connected to, often follows a power law. In a power law distribution, the histogram looks like a hockey stick. In other words, a large majority of nodes have less-than-average degrees and a much smaller subset of nodes have much higher degree than the mode or average degree across the set of nodes. (Watts, 2004, p. 250); (Barabási & Albert, 1999). By analogy, imagine a histogram of words in a given language. Some words will have extremely high use frequencies while others are specialized and therefore rare. The word frequency histogram will thus take on a power law shape. A further feature of such networks is that they are often robust to failure conditions modeled by removing edges from a graph, in part because there are many redundant paths through which information can travel. (Watts, 2004, pp. 261–262); (Dodds et al., 2003). Imagine an electrical grid with several different ways to supply power to a city. That grid is more robust to failure than one which depends on a single channel.

Finally, in part because they are robust to failure, small world networks are especially effective at quickly spreading information at distance. (Watts & Strogatz, 1998, p. 442). In fact, short average path length from node to node is usually the defining feature of a small world network, hence its name. Human social networks have this feature: Most people are usually friends of friends of friends of each other. In technical terms, many "real-world networks" have the features of small world networks since they are neither randomly generated nor completely ordered as a result of the vagaries of human interaction and the absence of hierarchically imposed ordering principles. (Watts & Strogatz, 1998, p.

[38] Watts and Strogatz's original paper created a small world network by randomly rewiring a ring lattice, producing the characteristic features as random rewiring produces a steep decline in average path length but negligible changes in clustering phenomena. (Watts & Strogatz, 1998, p. 440).

440).[39] The absence of "small worldness" would be cause to question the relative efficiency of information transfer within a network.

§5.5 Data Protection Epistemic Communities

To study the architecture of the network through which data protection outcomes might diffuse, this chapter foregrounds a subset of actors within the data protection "epistemic community." Haas defines an epistemic community as "a network of professionals with recognized expertise and competence in a particular domain and an authoritative claim to policy-relevant knowledge with that domain or issue-area." (Haas, 1992, p. 3). Such communities have shared normative and causal beliefs as well as shared notions of what constitutes acceptable evidence and what types of policy actions will solve the set of problems within the issue-area. (Haas, 1992, p. 3); (Bianchi, 2019). Haas adds, "epistemic communities are channels through which new ideas circulate from societies to governments as well as from country to country." (Haas, 1992, p. 27). And when the epistemic communities' expertise suggests that policy coordination is an optimal outcome, "epistemic influence" may lead to policy convergence, or analogously rule convergence. (Haas, 1992, p. 30); (Palladino, 2021).

Epistemic communities are a prime location for the study of data protection legal diffusion for several reasons. First and foremost, under conditions of uncertainty where the "correct" policy outcome is unclear—in this case what data protection legal rules to adopt—such communities "influence state interests either by directly identifying them for decision makers or by illuminating the salient dimensions of an issue." (Haas, 1992, p. 4). Those dynamics are present in the data protection space where uncertainty driven by the emergence of new technology and surveillance capitalist business models has placed many governments in a position where they have to turn to experts, many of whom belong to a transnational epistemic community. (Richie, 2010); (Bennett & Raab, 2006); (Bennett, 1997a). The structure of that community is therefore a theoretically fruitful place to begin to search for Lazer's architecture. Second, there is a rich literature on epistemic communities and legal diffusion to ground research. Thus, scholars have explored the nature of social relationships in the data protection space and how they influence legislation, (C. Dunlop, 2013; Palladino, 2021; Richie, 2010), and even the choices of private firms, (Cranor, 2002), or government agencies, (Fiodorova, 2018). Scholars have even recognized

[39] In the real world, the mechanisms which produce and grow small world networks are varied. For example, small world networks can come into existence when a social node emerges as particular well-connected. (Barabási & Albert, 1999). A further distinction between artificial and real-world small world networks is that in real networks the power law distribution sometimes has a sharp cutoff due to the fact that nodes age and adding marginal edges after a certain point is too costly—for example, humans have capacity for only so many social connections. (Amaral et al., 2000).

that the networks of actors themselves have a particular topology which influences the speed of diffusion and the normative value of the ideas that spread. (Lazer, 2005); (Lazer, 2006).

§5.6 Data Collection and Methods

One epistemic community of interest is the Privacy Law Scholars Conference, a conference run by a group of academics who share a commitment to the need for additional privacy protections. Members meet every year in a conference that brings together practitioners, researchers, and sponsors from a variety of countries in a multi-day forum. PLSC bills itself as "the premier academic conference [on privacy topics] in the world." And, helpfully, it publishes the membership of panel discussions for years going back to 2011. That provides a series of public edges (i.e., links between nodes) through which information about best practices could theoretically flow. It would, after all, be hard for members of the same paper panel discussion to avoid discussing ideas with one another. Using the public information as inputs, PLSC is an undirected graph with 814 unique individual nodes and 1709 unique edges suggesting shared membership on a panel. To capture the fact that paper panel ties become weaker over time, an exponential decay function, ($y = a(1 - r)^t$), where y is the edge weight, a is an arbitrary constant, 1 - r is the yearly rate of decay, and t is years since present, was applied to age the edges by year of formation. Information was manually compiled and coded.

Of course, another way to get at the relationships within the data protection epistemic community is to analyze the structure of knowledge within the available corpus of published scholarship. This chapter approaches that task in two ways. First, it manually compiled a citation network for the Top 20 papers in Google Scholar by citations per year for the Boolean search terms ["data protection & law"] and ["data privacy" & "law"]. Each unique citation in the resulting 40 papers (or in some cases books) comprises a node in the network TCN while the citation itself is linked via an edge with the paper in which it appears. TCN, or top citations network, is an undirected graph with 891 unique nodes and 994 unique edges. Second, the chapter manually compiled a citation network for the papers, books, and materials most heavily cited to construct the tripartite debate from Chapter 4. CDP, or comparative data protection, is an undirected graph with 620 unique nodes and 711 unique edges.

To assess the small world nature of each of the graphs, two statistics were computed: the small world coefficient sigma and the small world omega. Typically, small worldness is computed by determining the ratio $\sigma = \gamma/\lambda$ where $\gamma = C/Crand$ and $\lambda = L/Lrand$.[40] (Telesford et al., 2011, p. 368). This is referred to as the small world coefficient or small world sigma. Because the nature of a small

[40] Crand and Lrand refer to the average clustering coefficient and path length of a random network with the same number of nodes.

world graph is likely to create a small average path length but a much higher clustering coefficient, (Koniaris et al., 2014, p. 9), a graph is commonly classified as small-world if the small world sigma is greater than 1, which is the value expected for a completely random graph. (Humphries & Gurney, 2008).

The small world sigma has the downside of being overinclusive. There will be false positives because of small world sigma's inherent sensitivity to small variations in the clustering coefficient of a random network. (Telesford et al., 2011, p. 369). Another downside is that it gives "no sense of where along the spectrum of lattice to random," a given graph falls, that is how similar the network is to an ideal-type graph where every node is connected to every other node or instead to a randomly connected graph. (Telesford et al., 2011, p. 372). To account for these issues, an alternative metric is the small world omega. Small world omega is the difference between the ratio of the average path length of a random graph with the same number of nodes and the average path length of the graph under study and the ratio of the clustering coefficient of the graph under study and the clustering coefficient of a lattice graph with the same number of nodes. This is captured mathematically by the equation, $\omega = \left(\frac{Lrand}{L}\right) - \left(\frac{C}{Clatt}\right)$. (Telesford et al., 2011, p. 369).[41] The values of the small world omega are restricted to a range of -1 to 1 regardless of the network size; the closer the result is to 0 the more a graph exhibits the characteristics of a small world network.

A word of caution: While the sample graphs here provide different angles from which to view the social structure under study and the structure of knowledge within the policy community, they are incomplete. Only network traces that appear in textual artifacts were included. That means conversations and other such information exchange opportunities (e.g., information seeking on the Internet) that do not leave tangible traces are missing. Proprietary network information such as the membership of the IAPP, a transnational community of professionals in the data protection space, was also not included. And likewise, important social relationships such as those that arise from the Data Protection Commissioners meetings were also not included. In essence, only one data protection social network was captured, that of the PLSC. While there is reason to believe that other similar networks, such as the Privacy Law & Security Forum have similar structure (see Appendix B.1), the generalizability of the results here is debatable at best. Nevertheless, in exploring the "architecture" these graphs are steps in the right direction.

[41] The construction of a lattice graph using the algorithm developed by Sporns and Zwi, 2004, as in Telesford et al., was not undertaken here. Instead, a ring lattice, where each node was connected to the number of adjacent nodes equivalent to the average degree of the network under study, was employed for sake of simplicity.

§5.7 Results and Analysis

Results are broken down into two sections. First, Section 5.7.1 addresses the two small world metrics employed. None of the graphs display small world features. Second, Section 5.7.2. addresses the prevalence of siloing and of EU-dominant nodes within the network. Together these findings further undermine an inference that policy learning is taking place on a wide scale.

§5.7.1 The Absence of Small Worldness

The results of each metric for the three studied graphs are presented in Table 5.3. None of the graphs displayed small world features. Each graph has a small world coefficient (sigma) that is less than 1, indicating that the graphs do not display typical small world features compared to a random graph.[42] PLSC, for instance, has a small world coefficient that is significantly less than 1. Likewise, the small world omega for CDP and PLSC do not approach 0, the ideal value of a small world omega for a graph that is small world. (Koniaris et al., 2014). PLSC is very close to -1, which represents a completely random graph. CDP and TCN are much closer to lattice graph networks than random graph networks but neither is within a tenth of the small world value. In combination with the fact that each graph's small world coefficient is less than 1, it is safe to say that none of the graphs exhibit small world features.

	TCN	CDP	PLSC
Avg. Degree	2.231	2.294	4.199
Avg. Path Length	3.539	3.294	4.739
Avg. Clustering Coefficient	.367	.3	.634
Small World Coefficient (σ)	.597	.520	.105
Small World Omega (ω)	.111	.216	-.756

Table 5.3. Small World and Small World Omega Coefficients.

Interestingly, the knowledge networks TCN and CDP appear to be closer to demonstrating small worldness than the social relationships network PLSC. This is surprising as many social networks are small world. One speculative reason why this might be the case is that privacy law scholarship often has subareas of inquiry that do not relate to one another. For example, one group of scholars might research and attend panels on the economic analysis of how much individuals value privacy while another group

[42] In this case, Erdös-Rényi random graphs with an edge probability of 0.3 were employed. Lattice graphs were constructed using a rectangle lattice with the same number of nodes as each graph and the same average degree as the real-world graph under comparison.

might be focused on the case law of a specific country. This would have the effect of creating smaller components, discrete connected subgraphs unconnected to larger subgraphs, which could inhibit small world phenomenon.

§5.7.2 Siloing and the Centrality of Nodes with EU-Philic Views

Setting aside the lack of "small worldness," two other features of the graphs undermine the inference of policy learning as a mechanism of data protection diffusion. First, each of the graphs displays a large amount of siloing, the prevalence of "dead-end" nodes connected only to one other node. One can see the evidence of siloing in Figure 5.5, which is a graph of all the CDP and TCN nodes and edges in one location. It is therefore a larger n estimation of the nature of the knowledge network within the field. The average degree is 2.37 but approximately 95 percent of the nodes have a degree less than or equal to two. 88% of nodes have a degree of 1, meaning they are dead-ends in the network. Information which originates in such nodes has only one path to the rest of the network and information which originates elsewhere in the network must flow through that path to reach the dead-end node. That means that information flow is likely to be cumbersome and outcomes of legislation elsewhere, a crucial component of learning, are likely to diffuse slowly, if at all, on this graph.

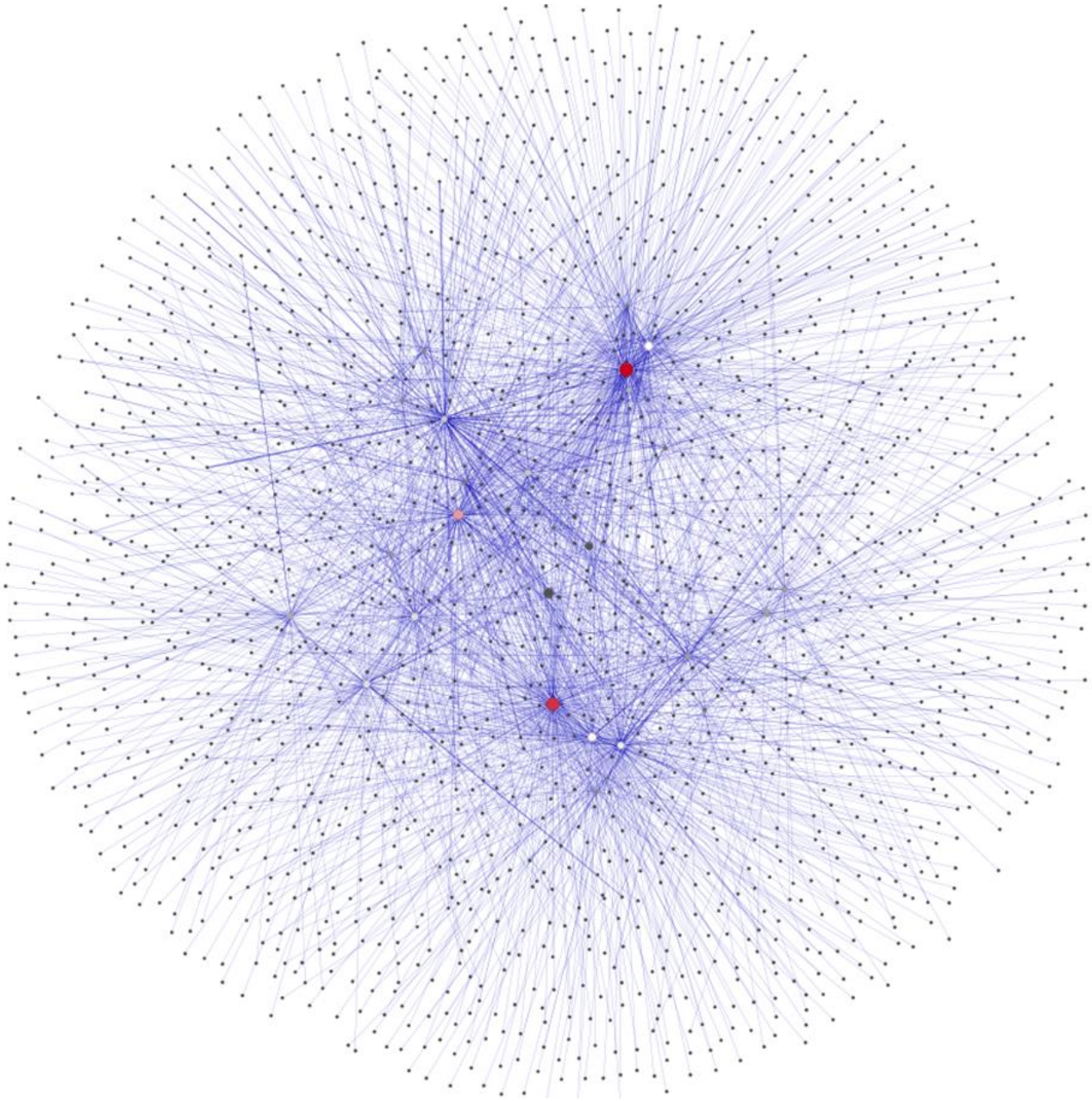


Figure 5.5. Unfiltered and combined TCN and CDP network graph. Edges indicate undirected citation. Nodes indicate individual papers. Layout by Fruchterman Reingold algorithm.

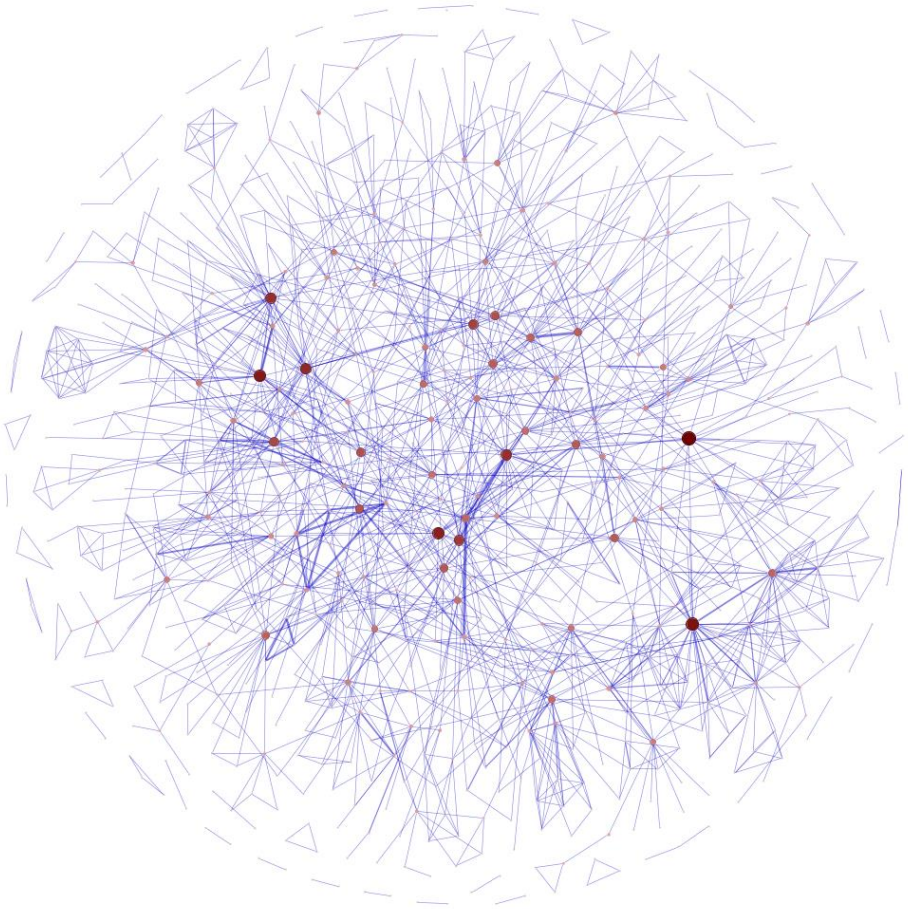


Figure 5.6. PLSC Graph with node color indicating degree and edge color indicating edge weight (darker blue indicates more recent links). Node color towards the red end of the spectrum indicates higher degree. Layout by Fruchterman Reingold algorithm.

PLSC, the graph of social interactions at the Privacy Law Scholars Conference represented by panel participation, displays similar features. In fact, many of the panels are isolates, unconnected to other parts of the graph (see Fig. 5.6). It remains to investigate the largest component in Figure 5.6. But the aged weight of the edges, as indicated by darker areas of blue, suggest visually that the central architecture is being driven by the most recent edges, that is by panel participation in the most recent iterations of the conference. Perhaps this is indicative of some degree of calcification in collaboration patterns over time since 2011. Further evidence of isolates is present in the TNC graph. It clearly shows the fact that one of the papers from the Google Scholar data set failed to have any shared citations. It is perhaps a good example of the fact that the literature on data protection often falls into a variety of buckets which do not always communicate with each other. Moreover, the nodes that link the papers

together, i.e., those nodes which fall within the red outline in Figure 5.7, nodes which could arguably form a shared canon, are sparse.

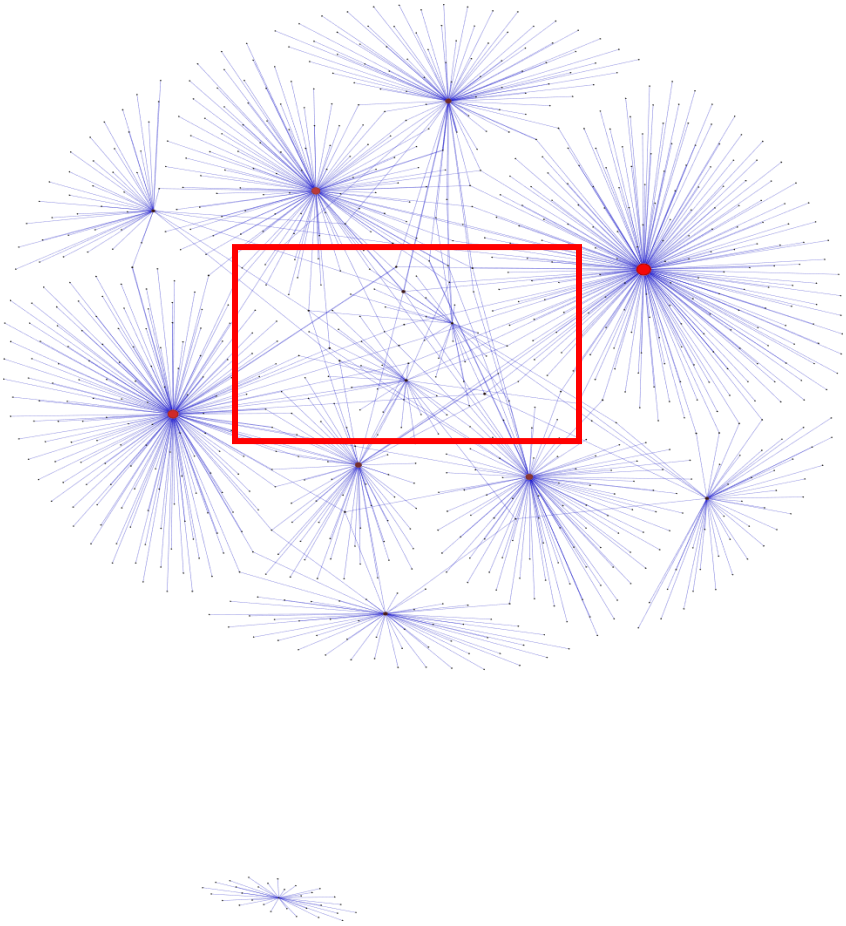


Figure 5.7. TCN filtered for the Top 10 Nodes by Citations/Year. Larger node size indicates larger degree. Layout by Fruchterman Reingold algorithm. Nodes that link the network together are outlined by red rectangle.

Together Figures 5.5, 5.6, and 5.7 tend to suggest that both social and knowledge networks in the data protection space are somewhat fragmented. Scholars tend to work on their own areas which cite the same papers without interacting much across subfields which touch on data protection issues. That is not auspicious for policy learning, which depends on such epistemic communities efficiently synthesizing and communicating diffuse information about policy outcomes.

A further ill-omen for policy learning is the fact that central nodes in both CDP and TCN express EU-centric viewpoints, suggesting that information about comparative policy outcomes is less likely to flow through such nodes. In Figure 5.8, for example, crucial nodes—as defined by their statistical degree centrality, or the relative number of other nodes to which they are connected compared to other nodes—are nodes which espouse an EU-model-centric view.

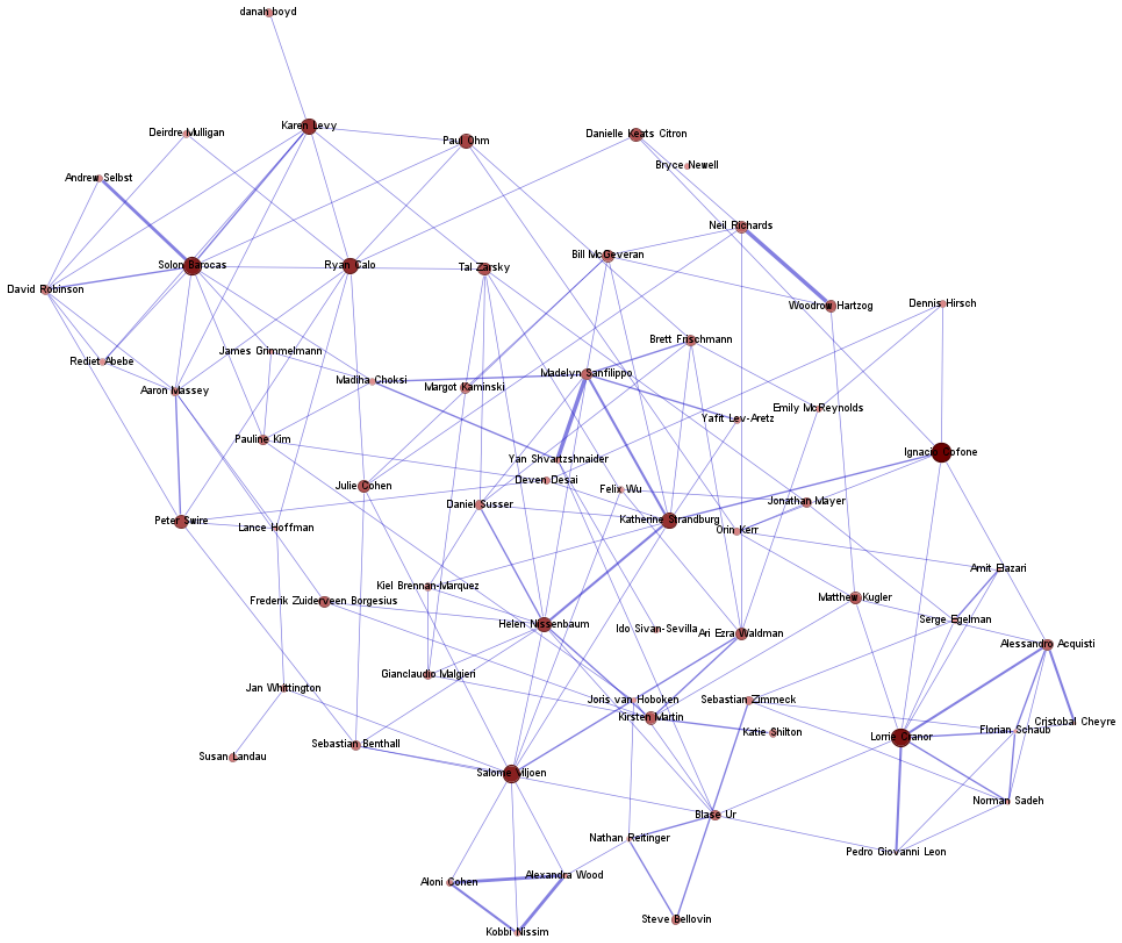


Figure 5.8. PLSC graph filtered for nodes with a degree equal to or greater than 13. Larger node size indicates larger degree. Layout by Fruchterman Reingold algorithm.

Such nodes include Ignacio Cofone, Katherine Strandburg, Salome Viljoen, Helen Nissenbaum, Lorrie Cranor, and Solon Barocas. This may be the result of these individuals' relative prestige in the community, and thus the desirability of recruiting them for a paper panel. Each expresses a preference for the EU model in hortatory terms. Nissenbaum, for example, favors the opt-in approach to processing and the shifting of responsibility to the state from the individual that the EU model fosters. (Nissenbaum, 2017, p. 594). Hiding little, Strandburg, Nissenbaum, and Viljoen refer to American sectoral laws pejoratively as a "great regulatory dodge." (Strandburg et al., 2024). And in her recent work, Viljoen, while stopping short of advocating for the EU model, concedes that its dignitarian approach accounts

well for the relational nature of the information ecosystem. (Viljoen, 2021, p. 646). Cofone's work on privacy harms and data ownership is a further example. In a 2021 article, for instance, Cofone was highly critical of the American model for "underutilization" of use restriction principles that might address what he refers to as the moral hazard problem of data use. (Cofone, 2021, pp. 567–571). Of course, such work is not incapable of presenting comparative outcomes or seeing flaws in the EU approach (e.g., (Cofone, 2023, p. 96), but the general message that the EU model is preferable provides reason to think that at least some of the most important nodes in the graph have reason to reify EU approaches to the extent it corresponds with their scholarly views. That may not be the ideal condition for policy learning because it suggests that certain outcomes that do not correspond to that majority view might be (unconsciously) suppressed.

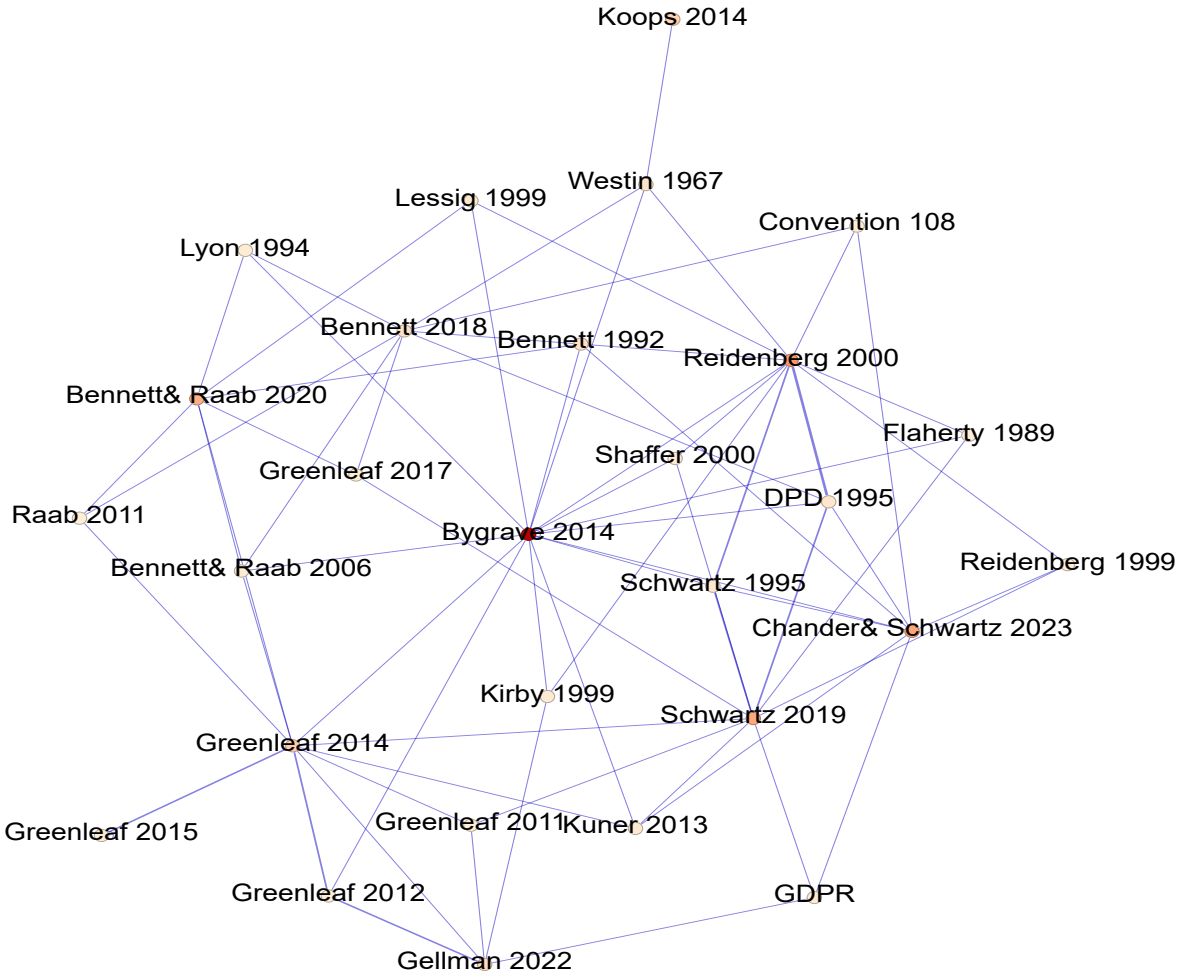


Figure 5.9. CDP filtered for nodes with a degree greater than 3. Layout by Force-Atlas 2 algorithm.

Turning to CDP, a similar phenomenon appears: Nodes marked by darker colors of red indicative of high centrality, express an EU-centric, even idealist view (see Fig. 5.9). Such nodes include Schwartz

2019 and Chander and Schwartz 2023. Schwartz, illustratively, adheres to the idea that EU data protection models rightfully have achieved “success in the marketplace of ideas.” (Schwartz, 2019, p. 818). Chander and Schwartz write that the only way out of the privacy thicket is to create a new global treaty whose norms would be developed “within an expert institution, such as the Global Privacy Assembly,” norms which would presumably be European in thrust. (Chander & Schwartz, 2023, p. 52). Of course, other important nodes like Bygrave 2014 are closer to the neutral comparison of outcomes that learning benefits from.

Interestingly, what Figure 5.9 may also show is that criticism of the EU-model is relegated to the periphery of important epistemic communities. Thus, it appears that Koops 2014, a paper highly critical of the EU model, appears as a peripheral node in the graph. Perhaps that is indicative of a consensus in the literature on the preferability of the EU template on roughly idealist grounds. That again undermines an inference that policy learning is occurring.

§5.8 Conclusion

Based on the results presented, including evidence from network analysis and comparison to known situations of learning, the likelihood of policy learning as a driver of data protection convergence is low. Policy learning, broadly understood as legislators incorporating outcomes from other jurisdictions’ legislation into their own decision-making, requires a flow of information regarding the impacts of other policies. But here, patterns in data protection align more closely with imitation, as reflected by high levels of textual similarity across jurisdictions and limited time intervals between legislative adoptions. Additionally, network analyses reveal that the knowledge-sharing infrastructure within the data protection community lacks the "small world" features that would facilitate the spread of policy outcomes, further weakening the case for learning. Instead, these networks are highly siloed, with many nodes having limited or no connectivity, curbing their ability to share policy outcomes effectively. Finally, the analysis highlighted the EU-centric orientation of central nodes in data protection networks, which suggests the possibility of a skewed flow of information that does not prioritize comparative outcome-based learning. Rather than providing a balanced view of various models’ effectiveness, these central nodes may advocate for the EU model, reinforcing a homogeneous perspective that limits the diversity of information within the network.

Of course, the conclusions here require qualification in some respects. First, they do not foreclose the possibility of learning in a specific set of cases. Second, the quantitative and network analysis is in some ways second-best evidence of a lack of learning. What would go a long way towards discrediting learning as a mechanism with additional finality would be admissions of imitation or copying

by rule choosers. Yet, to the extent such evidence is unlikely to be forthcoming or public, the quantitative approach taken here may be the best substitute. Furthermore, the conclusions are theoretically cabined to a limited set of legal diffusion curves and network graphs. Expansion of the data set would be a worthwhile development of the ideas presented.

Chapter VI

Explaining The Unexpected Persistence of Grouping: The Plausible Role of Domestic Law

One of the core findings in Chapter 3 was that grouping dynamics have gradually begun to surpass, if not entirely supplant, general textual convergence as a salient feature of global data protection monoculture. Even as countries' legislation has gotten textually more similar over time, the textual distances between grouped models seem to decline at a faster rate than the overall range of textual distances. Part II of Chapter 3 demonstrated that these convergences exist in part because EU ideas appear to be implemented in shared ways across jurisdictions within a textual grouping. And yet, these groupings are meaningfully different from one another as a statistical matter.

The persistence of groupings is somewhat perplexing. On balance, there would presumably be increased benefits from the network effect of a unitary model. Later-adopting countries would be able to "free-ride" on the model developed in the earlier-adopting country. (Garoupa & Ogus, 2006). And there are relatively few information barriers to converging on a single model. As comparative legal scholars recognize, "actors in all countries have access to the laws of other countries through the internet" and other means of information diffusion. (Kalantry, 2020, p. 58). There is also some evidence of attempts at formalized harmonization across regions. For example, Belli and Doneda document ongoing efforts to converge around a single model in BRICS nations. (2022). Why, given so many reasons to converge, do groupings persist? To recall the Queen Model, why and for what reasons do the tines of the queen's crown appear instead of a converging towards a point? (See Fig. 6.1).

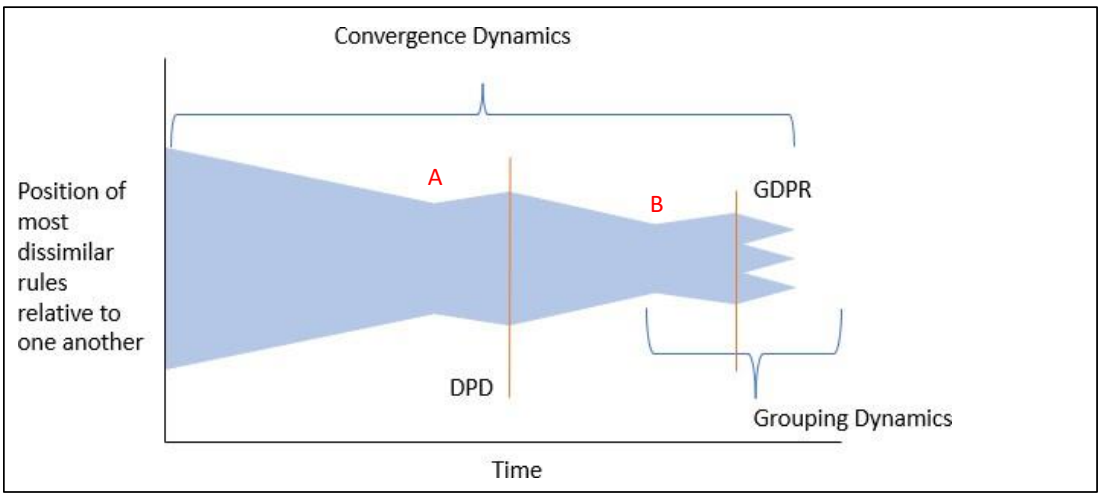


Figure 6.1: The "Queen" Visual Model of Convergence Dynamics.

Using the comparative law literature on legal transplants and elaborating on some of the findings in Chapter 3 regarding the characteristics of grouped jurisdictions, this chapter creates a list of plausible barriers to the further coalescence of grouped jurisdictions in Section 6.1. These include political, economic, and social factors such as the absence of shared languages or legal origins, or divergence in political power arrangements and economic characteristics. They also include the possibility that groupings are simply distinct in time, rendering further coalescence impossible because one grouping's model is or is perceived as too old to be adopted by a later grouping. It then shows that those factors are not plausible explanations for differences between groups.

Next, in Section 6.2, it employs the natural experiment presented by the adoption of data protection in Gabon, the American state of Montana, Tanzania, and Viet Nam (the "GMTV cases") to bolster that argument. The GMTV rules all were adopted within a year of each other, are single-party dominant states, have roughly similar economic indicators, and in some cases share legal origins and languages. The presence of meaningful textual divergence between them, in essence, the fact that they belong to different groupings, is thus strong evidence that factors which they all share (e.g., relatively low internet penetration) or share in part (e.g., English as a legal language) are neither necessary nor sufficient for groups to remain textually and legally different from one another. In other words, a factor such as the dominance of a single party does not result in a jurisdiction's text belonging to the same group as other grouped texts which also have a dominant single party and thus can likely be ruled out as a factor driving the difference between groups. The GMTV cases also provide strong evidence that countries are not simply looking to the adopters which are nearest in time to themselves when they create data protection rules. And while the GMTV cases cannot rule out the possibility that geography is a factor in continued textual differences across groups, there are strong reasons to think that geography is not a uniquely determinative factor. These include the ease of information transfer related to legal content over large distances and the empirical fact that the groupings beyond the GMTV cases are not always regionally-based.

The GMTV cases do, however, support the idea that preexisting domestic legal arrangements are shared across countries within groups but not between groups, meaning dissimilarities in domestic law may help explain continued textual distance across groups. As a result, one argument this chapter explores in Section 6.3—and about which future research is required—is that domestic law is a parameter which constrains the choice as to which foreign data protection model to adopt, ensuring that groups do not coalesce despite economic incentives to do so or shared geographic and socio-cultural factors. Domestic law could constrain choice in two ways. First, the rules governing the legal

requirements a data protection regime must satisfy could limit the transplantability of a foreign model. Second, the rules governing how law is interpreted might limit the translatability of certain models, further constraining choice.

Such limitations on the choices available to domestic actors would exacerbate the downsides of the global data protection monoculture. They ensure that even if policy learning is occurring (which is itself doubtful, as Chapter 5 shows), there are limits on the degree to which the best model can be implemented domestically if it conflicts with existing domestic norms and rules. That would be a further reason to doubt the normative value of rule monoculture.[43]

§6.1 A Range of Explanations for the Queen's Tines: From CDP to Comparative Law

Plausible explanations for the queen's tines can be drawn from the tripartite dialectic surveyed in Chapter 4 and from the comparative law literature. Unfortunately, the explanations these literatures generate do not have strong explanatory power as to the persistence of textual differences. Instead, the cultural, practical, historical, and economic factors they point to are much better at predicting which countries group than at explaining why there are as many groups as there are. In other words, they do not explain why there are multiple tines instead of one tine.

One preliminary question is whether grouped jurisdictions' data protection regimes are in fact genuinely distinct considering the reasons for them not to be. A critic might, for instance, argue that the rules only appear distinct when broken out into the separate time brackets employed in Chapter 3. And, pointing to Figure 3.15's heatmap of recent rules, a critic might add: isn't it true that convergence has reappeared as a dynamic? That would be the wrong approach.

Persistent textual diversity is the norm among jurisdictions that exhibit grouping. For the purposes of this chapter, jurisdictions are defined as exhibiting grouping if they are a member of a dyad with a Jaccard Distance of less than 80% of the average distance in their subset. Visually, these are the dyads that stand out from their peers as having a darker coloration towards the red end of the spectrum. Excluding subnational jurisdictions and any jurisdictions adopting prior to 1991,[44] roughly 36 sovereign states and overseas protectorates exhibit grouping based on the heatmapping performed in Chapter 3. Across dyads within the group of 36, no dyad has a Jaccard Distance less than .34 (Antigua & Barbuda and St. Kitts & Nevis), which is consistent with known copying and the average distance is .742,

[43] More broadly, the argument that domestic law creates path dependent dynamics by constraining the choice between foreign models is an important addition to comparative legal globalization debates. What it shows is that there may be an upper limit on the degree to which jurisdictions can become more alike one another that is inherent in law.

[44] This was for two reasons. First, it helps ensure that changes in the economic or cultural factors surveyed between the point of adoption and the present do not influence the results. Second, it puts the analytical focus on countries who have adopted laws during periods where the heatmapping reveals grouping dynamics.

consistent with instances of known legislative influence but not known copying (see Section 3.3). So, at base, the 36 jurisdictions in question do not have identical legislative texts, or even legislative texts that are relatively statistically similar, compared to other dyads in the set or benchmark examples of copied laws. What the empirical results of textual analysis instead suggest is that the groupings remain distinct from each other.

Why might that be the case? As in Chapter 4, one area to begin exploration is CDP theories of convergence. Unfortunately, CDP predicts relatively little remaining diversity in text. That makes sense. The CDP literature is primarily concerned with how to explain the general convergence in data protection law across jurisdictions. To many CDP scholars, EU models of data protection like GDPR "have acted as a coercive force, with significant economic consequences for those businesses that rely upon the unimpeded international flow of personal information, and on those governments that wish to protect their domestic industries from the possible consequences of non-compliance." (Bennett, 2018, p. 244). As the coercive force of such models grows "as more countries [join] the data protection 'club'," (2018, p. 240), coercive forces grow over time, presaging nearly complete convergence on the EU model. In economic terms, since the EU has a combination of market power, regulatory capacity, political will for stringent regulations, and a desire to regulate products or practices which it is costly to produce or carry out to different levels of privacy protection, convergence on the rule selected in the EU is almost inevitable, according to this realist view. (Bradford, 2012, 2015, 2020). Thus, answers to why diversity persists and why grouping occurs, if offered, are mere byproducts of theorizing designed to answer a separate question.

Of course, the degree of convergence CDP scholars predict relates to their views about whether diffusion is occurring via realist, idealist, or incrementalist means. Not all CDP scholars see complete convergence as the necessary endpoint of the diffusion of data protection models and ideas. For example, Schwartz argues that one reason why the diffusion of the EU model is both relatively free from coercive pressures and normatively good is that the EU demonstrates "considerable flexibility" when making adequacy determinations. (Schwartz, 2019, p. 786). Thus, jurisdictions can participate in global data trade while partaking of the portions of the EU model that best fit local circumstances. That view implies a degree of persistent dissimilarity in data protection schemes around the world. And, of course, most CDP scholars recognize that even if there is a rapid convergence on the EU model, there remains diversity in the scope of covered entities, implementation methods, and enforcement effectiveness. (Bennett & Raab, 2020, p. 451). But even CDP idealists broadly agree that convergence is likely and is in fact occurring. So again, the question becomes why are there not fewer groupings over time? For the

idealist, it is a question of why more countries have not rationally determined the EU text is best and subsequently adopted it. For the realist and the incrementalist, the question is why more countries have not felt pressure to select the EU text or realized that EU-based text is the only practical choice. Why textual grouping might develop is orthogonal to those questions.

A more productive place to look for explanations is the comparative law literature—a literature replete with examples of barriers to the movement of legal ideas across borders. Those barriers provide plausible explanations for the persistence of diversity because they create friction in the movement of ideas which may prevent complete convergence on a single model. Much of this literature deals with legal transplants. (Watson, 1993). Conceptually, the metaphor of a "legal transplant" is meant to evoke the image of legal rules from abroad as organic entities that a jurisdiction can accept, reject, or adapt to, not unlike a recipient body accepting or rejecting a transplanted organ. (Goldbach, 2019, p. 584; Teubner, 1998).[45] Transplantation encompasses the idea that the movement of legal rules and ideas across borders is a process that unfolds over time, driven by actors with a range of different incentives, rather than a rote process of copying and pasting. (Cohn, 2010). That makes it highly relevant to generating hypotheses related to grouping dynamics. Gradually, comparative law has coalesced around the position that for a law to truly move across borders, it must "'fit' within the institutional context in which justice is administered in a particular country. (Goldbach, 2019, p. 592; Twining, 2005).[46] Thus, a borrowed legal concept or rule can "fail" even if it remains on the books for an extended period of time. (Siems, 2018, p. 109). "Failure" can take a variety of forms including lack of enforcement, incoherence with existing norms, or incomplete adoption in local case law. (Berkowitz et al., 2003, p. 189). In contrast, success is often defined in the comparative law literature as whether the transplanted law has "adapted … to local conditions." (Berkowitz et al., 2003, p. 168).[47]

The reasons for the "failure" of a legal transplant provide a jumping off point for theorizing about the continued cross-group diversity in written data protection legislation. There are broadly two categories of thought as to why transplants might fail. The first is that domestic legal cultures prevent the reception of a transplant, not unlike an immune system rejecting a transplanted organ. There is a

[45] While the concept of what is and is not a legal transplant, and even whether the terminology is appropriate compared to something like legal diffusion or legal translation, is a matter of debate, the field has settled on transplant as "a conceptual tool" to "study the movement of law from one jurisdiction or legal system to another." (Goldbach, 2019, p. 584).

[46] In early iterations, comparative legal scholarship treated laws as "distant from [] social reality." (Watson, 1995, p. 469). In Watson's view, for a law to be transplanted it needed only to be literally copied: "[The] recipient system does not require any real knowledge of the social, economic, geographical, and political context of the origin and growth of the original rule." (Watson, 1976, pp. 80–81). Thus, the "[t]ransplanting of laws," in Watson's view, "was easy." (Cairns, 2013, p. 640). That view has not been widely accepted.

[47] Success may also have a broader valence in that it can be defined as whether the transplanted law has "the same impact in the reception country" as it did in the origin country. (Kalantry, 2020, p. 54).

strong and a weak version of this argument. In the strong version, there is no such thing as a successful transplant. The "similarity of solutions to given socio-legal problems across legal cultures" is a "deception" on this view. (Legrand, 1997, p. 124). The copying of a formulation from another jurisdiction is simply "convenient," rather than indicative of any underlying similarity in the legal systems, legal meaning, or legal practice. (Legrand, 1997, p. 121). For proponents of the strong view, "[a]t best, what can be displaced from one jurisdiction to another is, literally, a *meaningless* form of words." (1997, p. 120). The second, "weaker" position is to suggest that there are inevitably local socio-political and legal understandings that alter the meaning of a borrowed rule (in what Graziadei refers to as a process of "mediated action" (Graziadei, 2009)), but to accept that such similar rules qualify as transplants (Kahn-Freund, 1974) or at least legal irritants which produce legal effects in the recipient system (Teubner, 1998). Ginsburg's analysis of judicial review in South Korea and Taiwan falls within this tradition. (Ginsburg, 2002). Grande's point that occasionally the transplantation of foreign legal ideas results in an "inoculation effect," whereby domestic local legal systems respond to successful transplantation by pushing back on further transplantation and reifying preexisting norms, also fits within the weak tradition. (Grande, 2016). Both the weak and the strong version of the legal culture thesis suggest that the absence of a shared legal culture is a plausible explanation for why groupings remain distinct. Texts are simply not translatable across different legal-cultural norms so draftspeople do not try to import them. There is some empirical evidence for that view as well. Using a large data set of transferred rules, Bradford et al. conclude that the legal family to which a jurisdiction belongs is "an important predictor of legal substance in well-established legal regimes," though they note it "does little to explain substantive variation in more recent areas of law." (Bradford et al., 2021, p. 207).

The second category of thought as to why a transplant might "fail" (or never occur in the first place) is to suggest that there are practical limitations on transplantation. Such logistical barriers include the costs associated with writing and passing new legislation and of translating legislation originated far away or in a different language. Legal practice itself may present a barrier to the extent there are costs associated with updating understanding or domestic constituencies (like lawyers) who benefit from the status quo. (J. A. Miller, 2003, p. 855). Garoupa and Ogus acknowledge these costs in arguing that sometimes convergence on a single legal model (which they refer to as unification) is slowed by the fact that jurisdictions free-ride on the efforts of other countries to modify their laws. (Garoupa & Ogus, 2006, p. 359). These practical barriers also include geography and economics. Bradford, Chilton, and Linos find that a country's economic size is "consistently positively correlated with continued legal convergence and that a country's distance from the European Union is consistently negatively correlated with continued

convergence" in the area of antitrust rules. (Bradford et al., 2023, p. 35). The influence of political structures on the uptake of foreign rules has also been empirically explored. As one example, the degree of representation and integration in a past colonial legal system may be positively correlated with the receptivity of a post-colonial jurisdiction to law developed abroad. (Daniels et al., 2011).

It turns out that such factors, whether related to legal culture, practicalities, history, economics, or politics, are much better at predicting which countries group as opposed to explaining why there are as many groups as there are. By and large, the jurisdictions which appear in blocks of textual similarity are more dependent on their connections to global supply chains than their ungrouped peers, have smaller domestic economies, and relatively few people. On average, grouped jurisdictions have a higher percentage of GDP spent on imports (51.17%) according to World Bank data than the world average (which has fluctuated between roughly 25 and 29 percent in the last five years). Likewise, they tend to have negative current account balances per World Bank data. And they tend to be small, averaging roughly 17.28 million people per jurisdiction, and only 13.66 million if the Russian Federation is excluded. The average size of their economies is also relatively small, as the average GDP per capita across the jurisdictions is roughly $1,880 per person. For comparison, the world average is approximately $13,180 thousand per person, according to World Bank data from 2023. Furthermore, the grouped countries are heavily dependent on foreign direct investment, averaging roughly 6.9 billion U.S. dollars in FDI.

Collectively, this makes some sense. Those countries most dependent on trade and on domestic connections to the world economy would seem to benefit the most from adhering to a standardized data protection model which eases trade frictions across borders. (Garoupa & Ogus, 2006, p. 340). Likewise, it accords with the empirically-well-supported proposition that the likelihood of a given jurisdiction updating its laws to converge with a leading regulator in a "given policy area is" "proportional to the size of [its] economy and inversely proportional to [its] geographic distance" from the leading regulator. (Bradford et al., 2023).

There are also cultural and historical factors at play that distinguish grouped jurisdictions from ungrouped jurisdictions. Approximately 58 percent of the grouped countries share English as an official language, higher than the world average of roughly 40 percent. Many also have colonial histories (appx. 69.4 percent). Again, this makes some sense. English-as-official-language jurisdictions and those with colonial histories would seem well-equipped to borrow data protection legislative text from abroad compared to non-English-speaking jurisdictions without such a history. (Del Duca & Levasseur, 2010, p. 16; Kalantry, 2020, p. 56).

What these factors do not explain is the continued persistence of differences across the identified groups. As many grouped countries share such characteristics there is nothing to distinguish groups from one another based on these factors. Unsurprisingly, a multivariable linear regression assessing the effect of some of the variables discussed above, including FDI, GDP, colonial influences, government type, and account balance on average textual dissimilarity[48] within the 36 jurisdictions produces an R-squared of 0.098, indicating an extremely poorly fitted model. (See Figs. 6.2 and 6.3).

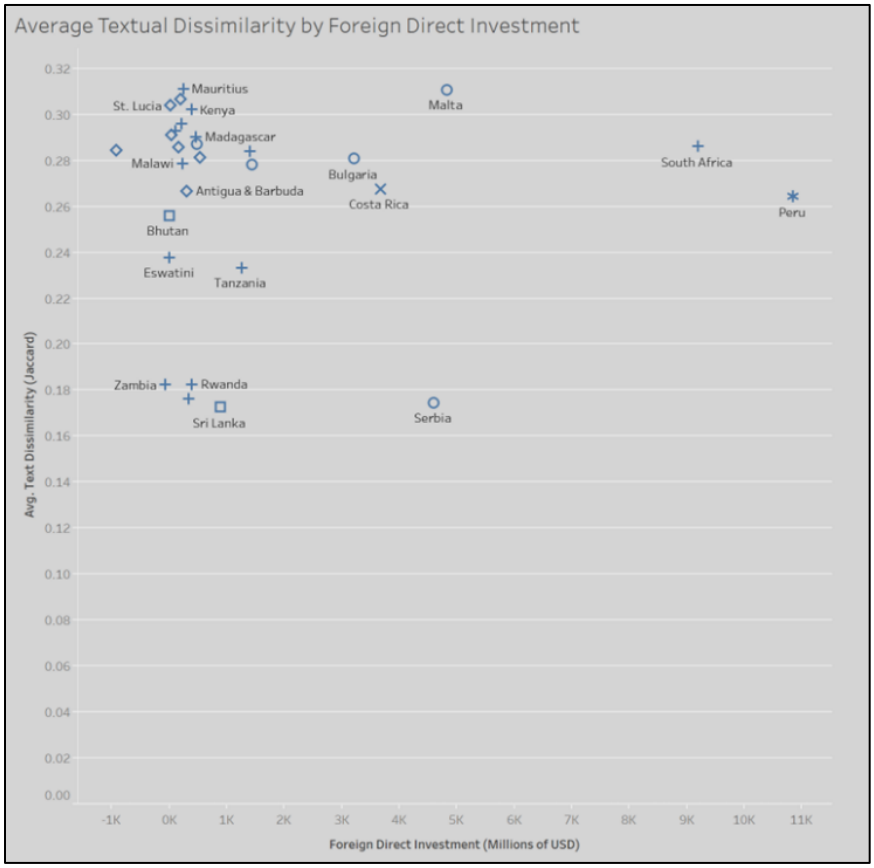


Figure 6.2: Average Textual Dissimilarity by Foreign Direct Investment. FDI Source = World Bank Data. Shared shape of country indicator denotes shared region. Outliers (e.g., Russia) excluded.

[48] Defined as $(\sum_{n=1}^{36} 1 - Jaccard\ Distance)/35$

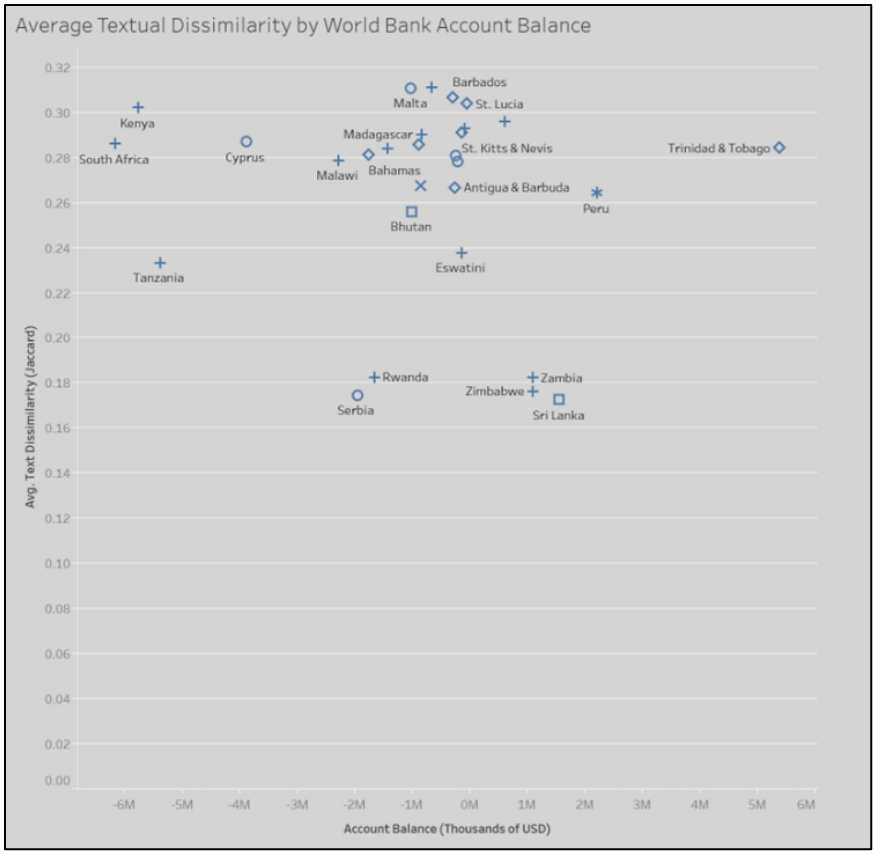


Figure 6.3: Average Textual Dissimilarity by World Bank Current Account Balance. Current Account Balance Source = World Bank. Shared shape of country indicator denotes shared region. Outliers (e.g., Russia) excluded.

As a further example of the inability of these factors to explain the persistence of distinctive groups, one can compare the groupings of Caribbean nations, particularly St. Kitts and Antigua and Barbuda, to those Southern African countries, particularly Zambia and Zimbabwe, identified as displaying grouping dynamics in Chapter 3. Each pair has a Cosine and Jaccard distance among the lowest of dyads in the set. All four countries share a Nineteenth Century colonial experience with British administration, English is a widely spoken legal language in each country, each country has a relatively small economy dependent on international trade, and a relatively low population compared to other regional players. And yet, rather than group together, the four countries form two distinct groups wherein each countries' rules are more like their regional neighbor than to the other two countries. This is unsurprising. As a practical matter, St. Kitts and Nevis and Antigua and Barbuda share many more links, both economic and legal, to each other than to Zambia and Zimbabwe. For instance, they share a Supreme Court (the Eastern Caribbean Supreme Court). The same is true for Zambia and Zimbabwe regarding their respective relationships with each other versus with St. Kitts and Nevis or Antigua and Barbuda. They share a land border. Both are members of the African Union and the Southern Africa Development

Community (which has circulated a model law on data protection). Another example is Costa Rica, Peru, and the Dominican Republic, which together form a block of textual similarity (see Fig. 3.13), in comparison to Spain. They share a language and legal origin yet remain textually distinct.

The obvious answer to the puzzle could be geography. Costa Rica, Peru, and the Dominican Republic are all much closer to each other than to Spain and the same is true of St. Kitts and Nevis and Antigua and Barbuda in comparison the Zambia and Zimbabwe. Ultimately though, high geographic distance does not appear to be a primary factor in explaining the continued distinctiveness of groups. Geography, of course, may influence the degree to which legal rules are shared and the cognitive and practical availability of models, with models developed closer being more accessible to rule choosers. But in a globalized world with relatively few information barriers geographic distance is unlikely to be a serious hurdle to borrowing from rules created elsewhere. Illustratively, one of the lowest Jaccard Distances is between Israel and the Bahamas (see Appendix A), and one of the most visually distinctive groupings in Figure 3.13 involves Bhutan and Malawi—not countries linked by geographic ties. Likewise, many of the groupings are relatively close to one another but distinct in terms of the laws to which they are most similar. Consider, for example, Kenya and Botswana which are closer to the Antigua and Barbuda group in the dendrogram in Figure 3.13 than they are to the grouping of Madagascar and Mauritania. Or to use an earlier example when the geographic spread of data protection rules was lower, 1970s Austrian rules are textually more similar to 1970s Danish rules than to 1970s German rules, despite Austria being closer to Germany than Denmark (see Fig. 3.9). So while geographic distance may remain a factor which creates additional friction in adopting rules from far away, (Bradford et al., 2023), generally speaking it seems to play a non-determinative role in the persistence of groups, creating some friction for convergence but ultimately being not necessary for groups to remain distinct.

Several difficulties remain in generalizing from anecdotal comparison or from the absence of statistically significant regressions based on commonly identified factors, primarily the fact that groupings adopted rules in different time periods, making the argument that Zimbabwe, for example, simply adopted the more recent rules from Zambia, rather than the older rules from St. Kitts or Antigua difficult to rebut. Furthermore, because each of the groupings share so many cultural and historical features—to use the island and East African examples, English as the official language and a colonial history related to England—it is hard to control for such factors, leading to an inability to discard the possibility that the variations in each's rules are due to variations in the degree to which they share such characteristics (e.g., the degree to which English is spoken in the legal system) or variations in the shared characteristics themselves (e.g., the nature of colonial administration in each jurisdiction). Analyzing

jurisdictions that adopted rules at the same time but do not share cultural, historical, or legal factors would help control for such explanations. One option is the GMTV cases surveyed in the next section.

§6.2 The GMTV Cases

Gabon, the American state of Montana, Tanzania, and Viet Nam (collectively GMTV) are fit to purpose. Each passed their recent legislation within several months of one of the others beginning with Tanzania in November of 2022 and finishing with Gabon in September 2023. Each is a member of a textual grouping.[49] Each has similar degrees of inequality with Gini coefficients ranging from 36 (Vietnam) to 43 (Montana) and relatively large percentages of its population living in rural areas, consequently decreasing the degree of internet penetration (see Fig. 6.4). At the same time, they do not share historical or cultural factors. Two of the group share English as a legal language (Montana and Tanzania) while two do not (Gabon and Viet Nam); two share a common law heritage (Montana and Tanzania) and two do not (Gabon and Viet Nam, which both have elements of French legal influence), while two are relatively wealthy (Gabon and Montana) and two are not (Tanzania and Viet Nam) judging by GDP/capita (see Fig. 6.4).[50]

And finally, two are relatively undemocratic (Gabon and Viet Nam) and two are relatively democratic (Montana and Tanzania), (see Fig. 6.4), though all share single-party dominance in their political systems. In turn, each passed rules while under one party control. Gabon was from 1968 to August 2023 a de facto one-party state under the control of the Parti Democratique Gabonais, essentially the political arm of the powerful Bongo family. On August 30, 2023, President Ali Bongo, after being elected to a third term under dubious circumstances, was deposed by the Gabonese military and Brice Nguema, a military officer and Bongo family relative, was appointed “chairman of the transition.” (Gbadamosi, 2023). That said, there are no signs of a transition to multi-party democracy. Gabon’s data protection law passed the National Assembly and Senate and received “Presidential approval” on the heels of the coup on September 7, 2023, though it is unclear whether the law was de facto in place prior to the coup.

[49] See the figures in Appendix B regarding Viet Nam. It is relatively statistically similar to several other East Asian jurisdictions, especially Taiwan (see Fig. B.3).

[50] Moreover, though the *Autorité pour la Protection des Données Personnelles et de la Vie Privé* (APDPVP) in Gabon is a member of the Global Privacy Assembly, for the most part each jurisdiction is shielded from the Global Privacy Enforcement Network and the Global Privacy Assembly, two prominent networks of data protection enforcement and idea coordination, increasing the likelihood domestic models are sheltered from international idea flow.



Figure 6.4. Internet Penetration Percentage by Economist Democracy Rating for 2023. Size of data point indicates GDP/Capita relative to the other data points.

In Montana opposition parties are relatively weak and control of legislative power is not competitive. Since 1990, the Republican Party, a right-leaning party in Montana's two-party system, has controlled the upper house of the legislature for all but eight years. The opposition Democratic Party has not had control of the Montana Senate since 2009. The Democratic Party has fared even worse in the lower house. While it attained rough equality of representation in several electoral cycles where the Montana House of Representatives was officially split, it has only attained outright control for two years since 1990, from 1991 to 1993. Currently, the Republican Party has a supermajority, meaning it can pass bills without any opposition party votes and over the veto of the Governor, who normally would possess the power to kill legislation. (Kimbel-Sannit, 2023). Montana's data protection law was passed May 19, 2023, while the supermajority was in place.

Tanzania and Viet Nam share single-party dominance with Gabon and Montana. From 1977 to 1992, Tanzania was officially a single-party state and even after 1992 the National Assembly, Tanzania's legislature, has been dominated by the Chama Cha Mapinduzi party, a political conglomeration of the

former dominant parties in Tanganyika and Zanzibar. In the last Presidential election, the CCM candidate, John Magufuli, won nearly 85% of the vote share. Like Gabon and Montana, Tanzania's legislative apparatus was in the hands of a single party when its data protection law was passed on November 27, 2022. (Ojwang, 2024).

Legislative power in Viet Nam, like the other three jurisdictions, is also in the hands of a single dominant party, in its case the Communist Party of Vietnam ("CPV"). It is the CPV National Congress, which meets every five years and sets overall policy via official Resolutions, that is generally responsible for the direction of legislation, which is usually proposed and drafted by Government Ministries staffed by CPV members and subsequently passed in the semi-professional National Assembly, whose members are almost all CPV members. (Greenleaf, 2014, p. 363). Officially, the omnibus data protection rules issued April 13, 2023, about a month into the presidency of Võ Văn Thụơng, came in the form of a Decree issued by the Vietnamese Government at the "request" of the Minister of Public Security and pursuant to several broad laws passed by the National Assembly throughout the 2000s.

The presence of shared socio-political and economic indicators as well as the fact that each jurisdiction's rules were passed within the same period renders those shared factors implausible explanations for textual divergences. As the next three subsections demonstrate, the rules in each of the four jurisdictions do in fact diverge along several continua. These include the revealed goal of the regimes, the applicability and scope of the rules, their processing principles, and the specific rules related to penalties and cross-border data transfer.

§6.2.1 Divergent Revealed Goals Through Text and Legal Meaning

All laws communicate something to somebody via their text. Data protection laws reveal much about the way draftspeople view the state as an actor within an information ecosystem. They thus also reveal much about the aspirations of the state and whether it wishes to act as an arbiter, referee, manager, or director in an information ecosystem. Textual comparison of the four rules reveals strong differences in state aspirations across several common elements of data protection schemes.[51] In broad brush strokes, the GMTV rules can be classified along a spectrum of breadth in geographic scope from Montana at the low end to Viet Nam at the upper end. Each rule further sets out a series of processing principles—illustratively, each shares the goal of linking processing with a specified purpose—but these principles differ in important respects. Where Gabon and Tanzania focus on fairness and lawfulness—that is, the idea that processing is illegal unless it fits within a defined category permitted by law—

[51] For the purposes of this subsection, the rules were coded across data protection continua using NVivo.

Montana and Viet Nam focus on ensuring transparency, security, and data integrity. As the following sections show, Viet Nam created rules that are broad but uncomplicated to comply with in theory, all the while making the data economy transparent to the state and dependent on reciprocal obligations. Gabon created a regime with extremely detailed, possibly hard to comply with rules, that have reasonably large penalties, all the while textually committing Gabon to rigorous enforcement and international principles. Tanzania created a preregistration scheme designed to bring data processing under the control of the central Union administration. And Montana created a scheme with minimal state interference and oversight of processing, instead outsourcing it to the consumer and her representatives.

The GMTV rules do share some trivial similarities, including reference to the familiar controller-processor-third party distinction. For example, in Viet Nam actors within the information ecosystem are divided into *Bên Kiểm soát và xử lý dữ liệu*, or data processor*, Bên Xử lý dữ liệu*, or data controller, *Bên thứ Ba*, or third parties, and *cá nhân*, or individuals. In Gabon, the terms are *responsables de traitement*, data controllers, *sous-traitant*, subcontractors/functionally data processors, and *tiers*, or third parties. But the distinctions in revealed goal, breadth, and principles trump these similarities. Such distinctions are canvassed and analyzed in the sections below starting with applicability before turning to obligations and penalties.

§6.2.2 Applicability and Processing Principles

Beginning with the broadest application, the Vietnamese decree applies to "Vietnamese agencies, organizations and individuals" without geographic limitation and even when they "operate in a foreign country," "foreign authorities, entities and individuals in Vietnam," and any "foreign agencies, organizations and individuals that directly process or are involved in processing personal data in Vietnam." (Article 1(2)(a)-(d)). It notes that the processing of personal information, defined as information which identifies an individual (Article 2(1)), by any of those entities should be subject to "state management." (Article 5). Tellingly, it does not lay out a series of lawful bases for processing, except to require that data subject consent be attained beforehand.[52] It requires consent to be present prior to processing unless processing is necessary to "protect the life and health of the data subject," (Article 17(1)), disclosure of data held privately is required by law (Article 17(2)), or processing is required to "serve operations by regulatory authorities," (Article 17(5)). Consent must be "written," "is not presumed" in electronic transactions, and "acceptance of the general conditions of use or sale" is

[52] Viet Nam defines consent as "an act" with which the data subject permits the processing of their data "in a clear, voluntary and affirmative manner." (Article 2(8)).

not sufficient. (Article 72). Once consent is obtained, and in the case of sensitive data, explicit notification of processing is given (Article 11(8)), processing personal data requires only that "personal data" be "appropriate for the scope and purposes of processing" (Article 3(4)), and that personal data is "protected and secured throughout the processing," (Article 3(6)). An ongoing log of impact assessments must also be maintained by data processors (Article 24). That log must be "always available" for "assessment by the Ministry of Public Security, which must receive a log outlining the nature of processing with "60 days" of the start of processing (Article 24(4)).

Vietnamese legislation does not address lawfulness, fairness, or transparency within the boundaries of the document. Instead, the goal appears to be to make the processing of personal data dependent on consent and, maybe even more importantly, transparent to the state; the notion that private processing and actions are subject to the view of the state pervades the rules. For example, according to Article 18, no consent is required when "agencies and organizations" make "audio and video recording" "in public places to protect national security, social order and safety." But Article 18 also officially requires that data subjects be made aware they are being recorded. The notion of public surveillance is to be made itself public.

Somewhat like the legislation in Viet Nam, Montana's legislation appears to be reasonably light-touch in processing principles, though Montana does not appear to share the goal of making processing transparent to the state and it diverges drastically in its scope of application. It applies only to entities that "conduct business in" the state or target "residents" with products or services which "control or process the personal data of not less than 50,000 consumers," unless the entity does so "solely for the purpose of completing a payment transaction," or "control[s] or process[es] the personal data of not less than 25,000 consumers and derive[s] more than 25% of gross revenue from the sale of personal data." (Section 3(1)-(2)). It does not apply to any elements of the state of Montana including boards, agencies, districts, commissions, and authorities, any nonprofits, institutions of higher education, national security associations, large financial institutions governed by federal rules, most health entities governed by federal rules, most health care and job safety information regulated at the federal level, information as to consumer creditworthiness, and several other categories (Section 4(1)-(2)). Moreover, its rules do not apply, for the most part, to deidentified data, (Section 10(2)), or publicly available data, rather than personal data, which it defines as information linked to an "individual," (Section 2(15)(a)-(b)). It also takes pains to note a variety of further exceptions and interpretive principles that limit its application, including to specifically note that none of its rules should be "construed" to limit "freedom of speech or freedom of the press" (Section 11(5)(a)(i)), and that they should not be "construed" to apply "to a

person's processing of personal data during the person's household activities," (Section 11(5)(b)). In sum, there may be more ways in which Montana's rules do not apply than do apply.

Where the rules apply, Montana requires processing to be limited to what is "adequate, relevant, and reasonably necessary in relation to the purposes for which the personal data is processed," but like Viet Nam it does not explicitly address issues of lawfulness or fairness. Montanan data processors are nevertheless required to provide "consumers with a reasonably accessible, clear, and meaningful privacy notice," (Section 7(5)), and must "establish" "reasonable administrative, technical and physical data practices to protect the confidentiality, integrity, and accessibility of personal data," (Section 7(1)(b)). Unlike Viet Nam, Montana's regime is not consent-based. The consent of a "consumer" need not be obtained prior to processing. Instead, the consumer has the right to "confirm whether a controller is processing" their data and "delete" that data as well as "opt out" of processing that has already begun in the case of "targeted advertising," the "sale of the consumer's personal data," and "profiling in furtherance of solely automated decisions." (Section 5(1)(a)-(e)).

Montana's goal appears to be to make processing transparent to the "consumer" in their role as a consumer. Pertinently, it defines "consumer" as "an individual who is a resident" of Montana except when an individual is acting in a "commercial or employment context." (Section 2(6)(a)-(b)). In other words, all residents of Montana are consumers unless they are employees. This is in contrast to the other three jurisdictions, which divide actors into the common data subject-data controller-data processor-third party set up. And unlike Gabon, Tanzania, and Viet Nam, Montana's regime seems not to be designed as an information governance scheme but rather as a response to specific problems within the information ecosystem, such as targeted advertising or "dark patterns," which it defines as user interfaces "designed or manipulated" to subvert "user autonomy, decision-making or choice." (Section 2(9)).

Gabon and Tanzania exist somewhere between Viet Nam and Montana on the spectrum of application breadth. Gabon is close to Viet Nam in that it applies its data protection rules to any entities which collect, process, store or use personal data—defined as information which identifies or could identify a natural person, (Article 6)—by automated or non-automated means "in any place where Gabonese law applies." (Article 4). That includes any processing which occurs using "means located in Gabonese territory" unless solely for "transit." (Article 4). Interestingly, Article 4 also specifically notes that it applies to any processing concerning data linked to "public security, defense, the investigation of criminal offenses or the security of the State, even if linked to an important economic interest of the state." It does not apply to processing "exclusively within" "domestic activities" or "temporary copies" of

personal information for “technical activities of transmission” or “transitory storage.” (Article 4). Where it applies, the Gabonese legislation notes that its goal is to “combat attacks” on “privacy” “caused by the collection, processing, transmission, storage and use of personal data.” (Article 2). It therefore aims to ensure that “data is collected and processed fairly and lawfully” for “specific, legitimate and non-inhumane purposes.” (Article 70). Such data should be “adequate” but “relevant” to the purposes for which it was collected and “not excessive with regard” to such purposes (Article 70).

Those principles mean that data processing must be based on “the consent of the person concerned” or “compliance with a legal obligation,” “safeguarding the privacy of the person concerned,” “the performance of a public service mission,” “the execution either of a contract” or “pre-contractual measures taken at the request of the data subject,” or the “achievement of the legitimate interest” of “the controller.” (Article 71). Notice must be given before personal information “relating to private life” is collected and the methods of collection must adhere to a set “standard.” (Article 180). In Gabon, it is incumbent on the processor to demonstrate that the data subject has consented to processing when consent is the lawful basis of processing. (Article 72). Gabon also requires consent to be present prior to processing unless processing is necessary to “protect the life and health of the data subject,” (Article 17(1)), disclosure of data held privately is required by law (Article 17(2)), or processing is required to “serve operations by regulatory authorities,” (Article 17(5)).

Tanzania’s data protection rules apply to any collection and processing of personal data by a controller “domiciled” in the United Republic of Tanzania or in any “territory where the laws” of Tanzania apply by “virtue of international public law” as well as any controllers not “domiciled” in Tanzania who process personal data in Tanzania unless the processing is for “mere transit.” (Part IV Subsection 22.1(a)-(c)). It does not apply to Zanzibar regarding “non-union matters,” which per Tanzanian law includes issues like local government, road maintenance, local agricultural regulation, small scale industry, prisons, and any other issue which does not appear in the First Schedule of the Tanzanian Constitution. Processing “may” be exempted if the appropriate authority determines that personal data was processed by the data subject for personal use, in accordance with a court order, “for the purpose of safeguarding national safety and security,” preventing crime, tax evasion, or the misappropriation of public funds, or for vetting a person for appointment to an official position. (Part IX Subsection 58.2).

The Tanzanian processing principles overlap some with the Gabonese principles. For example, data must be processed “lawfully, fairly and transparently,” “collected for explicit, specified and legitimate purposes and not further [processed] in a manner incompatible with those purposes.” (Part I, Subsection 5(a)-(b)). Data collected must be “adequate, relevant and limited to what is necessary” for

the collection purpose as well as "accurate and where necessary, kept up to date." (Subsection 5(b)-c)). But unlike in Gabon, data processed must be subject to "appropriate security" "against unauthorized or unlawful processing [or] loss." (Subsection 5(g)). And, in general, data may only be collected "directly from the data subject concerned" unless the data is publicly available or such direct collection is not "reasonably practicable." (Part IV Subsection 23.1).

Also, unlike Gabon or Viet Nam, Tanzania's rules are not partially consent-based, except around sensitive personal data which does require "prior written consent" before processing. (Part IV Subsection 30.1). Rather than create a list of lawful bases of processing with consent one basis like Gabon, or with consent as the only basis as in Viet Nam, or like Montana assume that processing is default lawful unless the consumer opts out, Tanzania has a registration scheme akin to the earliest forms of data protection rules in Scandinavia and Western Europe. Per Part III, Subsection 14.1, "[a] person shall not collect or process personal data without being registered as a data controller or a data processor." Only once the Tanzanian regulatory authority issues a "certificate of registration" may collection or processing begin. The certificate is only good for five years. (Part III Subsection 16.1). An approved "code of ethics or policy for personal data protection" is also required before processing or collection can begin. (Part IX Subsection 65.1-.2). In contrast, Gabon has a pre-authorization scheme only for certain categories of processing, like the "automated processing of genetic data," or the processing of data related to criminal offenses, or the automated processing of data "concerning assessments of people's social difficulties." (Article 81).

Unlike the goals revealed through the other three rule sets, Gabon's larger motivation may be to project legitimacy and sophistication. It has several interesting and unique aspects which possibly betray such an aim. For one thing, it does many things all at once and its rules deal with an incredible wealth of technological issues and are voluminously detailed, breaking down processing rules in at least 50 different discrete situations. Illustratively, its rules touch on data protection, data breach (Articles 142 through 147), and, unusually, many areas of AI regulation. For example, one goal of its data protection rules is "to frame Artificial Intelligence in such a way as to make it trustworthy, human-centered, ethically sustainable and inclusive." (Article 177). Furthermore, the rules create a set of very detailed rules relating to children. For example, data controllers and online service providers "must establish a clear distinction between marketing aimed at children and marketing intended for adults." (Article 197). Online service providers are also prohibited from offering "prizes [or] rewards" with the goal of inducing children to "stay on the site or take part in activities." (Article 199). Its definitions section is equally voluminous, defining unusual terms not normally included in a data protection rule-set definitions

section. These include “ad exchange platform,” “machine learning,” “CNAME Cloaking,” “digital workspace” and the entire field of “computer science.” Such terms are not used in the remainder of the legislation. Their inclusion is thus more symbolic than functional.

Gabon’s rules also take several opportunities to bind Gabon to international conventions unrelated to data protection such as the International Convention on the Rights of the Child (Article 188). They also require the Gabonese DPA to develop a unique Gabonese position in “international negotiations in the area of the protection of personal data and privacy.” (Article 8). Concern for Gabon’s international legitimacy bleeds through the text. Its rules are unusually explicit in the length to which they textually bind the state (e.g., Article 4). Finally, somewhat tellingly, it requires data controllers to comply with the rules “as soon as [the] law is published.” (Article 219).

§6.2.3 Obligations, Cross-border Rules, and Penalties

A further difference in the four jurisdictions relates to the obligations the data subject (or consumer in Montana’s case) has. Each of the regimes provides a list of rights to the data subject. But only Vietnam also creates a list of obligations the data subject has to the information community. These include, “protecting his/her own personal data,” “respect[ing] and protect[ing] others’ personal data,” and “fully and accurately providing his/her personal data when he/she consents to the processing.” (Article 10(1)-(4)). Vietnamese organizations also have obligations, particularly to report violations either committed by themselves or by others to the Ministry of Public Security. (Article 42). These ideas arguably evince a notion of the information ecosystem as a community to which each element has duties, a unique aspect of Vietnamese legislation compared to the other three cases.[53] At the opposite end of the spectrum, far from giving the consumer obligations, Montana relieves the consumer of any obligations at all, allowing her to offload much of the responsibility for exercising her rights to an “authorized agent.” (Section 6).

The GTV countries may reveal their ambitions to create information governance schemes (especially in comparison to Montana) even more through their penalty and cross-border provisions. The combination of large penalties for violation and the imposition of transborder data transfer limits maximizes the law-based leverage each country’s enforcement architecture can exert on regulated entities. For violating the rules in Viet Nam “agencies, organizations and individuals” may “be disciplined, or face administrative penalties or criminal prosecution.” (Article 4). The exact nature of such penalties

[53] This may reflect a cultural commitment to “Confucian” norms, norms which Ginsburg has charted in the legal contexts of several East Asian states (Ginsburg, 2002), and Greenleaf notes may be prevalent in Vietnamese information law (Greenleaf, 2014, p. 362).

remains unspecified. Gabon and Tanzania outline a specific set of strict liability penalties. Data subjects who suffer damage in Tanzania from the violation of its rules are due compensation per Part IV, Subsection 37.1 without a showing of a fault. "Obstruction" of the Tanzanian Personal Data Protection Commission ("PDPC") may result in criminal conviction with a prison term of up to two years or a fine of between one hundred thousand and five million Tanzanian shillings (appx. between $40 to $2,000 dollars). (Part VII Subsection 43(a)-(d)). Unlawful disclosure of personal data may result in fines of up to "twenty million shillings or to imprisonment for a term not exceeding ten years," or in the case of a "company or corporation" a fine of up to five billion shillings (appx. $2 million). (Part IX Subsection 6(a)-(b)). Unlawful destruction or deletion of personal data can also lead to a prison term and a lesser fine. Interestingly, "every officer" of a company who "knowingly and willfully" allows the offense is liable for the penalties. (Part IX Subsection 62). In Gabon, "obstruction" of the APDPVP, the data protection regulatory agency, can result in prison terms of up to a year and a fine of between one million and ten million West African francs (appx. between $1,700 and $17,000), significantly more money than the fines in Tanzania. The Gabonese APDPVP also has the power per Article 203 to order a regulatory fines of up to one hundred million CFA francs (appx. $170,000). Repeat offenses may result in fines of up to three hundred million CFA francs or, in the case of a company, 5% of yearly global turnover up to a limit of ninety-six million CFA francs. (Article 204).[54]

Each of the three countries also have unique international data transfer rules. Data controllers subject to Gabon's rules "may only transfer personal data to another State with the approval of the APDPVP," which by rule is required to permit transfer only when other states ensure "a sufficient level of protection of privacy and fundamental rights." (Article 171). An exception may be made where the transfer is subject to "guarantees," such as "contractual clauses or internal rules" which provide a "sufficient level of protection of privacy." (Article 173). Further exception may be made where the data subject consents to the transfer and it is "necessary" to "safeguard the life of that person," "to safeguard the public interest," or to conclude a contract in the interest of the data subject or to which the data subject consents. (Article 173). The Tanzanian cross-border rules are similar. The PDPC has the power to "prohibit the transfer of personal data to a place outside [Tanzania]." (Part V Subsection 31.1). And in any case, transfer may only be to a country "that provides for adequate data protection;" "the recipient [must] establish[] that the" transfer is necessary for the "public interest" or "pursuant to the lawful functions of a data controller" or that there is "no reason to assume" the data subject's "legitimate

[54] Why the global turnover provision has a lower upper limit than the general fines provision is left unstated.

interests might be prejudiced by the transfer." (Part V Subsection 31.2). Where the transferee state lacks "adequate" data protection rules, data may be transferred where the data subject consents to transfer, the data controller has specific authorization from the PDPC, the transfer is necessary for a contract between the data subject and the data controller, the transfer is necessary to protect the data subject's "legitimate interests," or is legally required on public interest grounds. (Part V Subsection 32.4(a)-(f)). At the same time, Tanzanian rules permit the PDPC to allow transfer subject to "adequate" "safeguards" such as "contractual clauses." (Part V Subsection 32.5).

Viet Nam, and to a greater extent Montana, is significantly more lenient regarding cross-border transfer of data, requiring only that the sender keep a log of the information sent and the conditions in which it was sent and notify the Minister of Public Security in advance. (Article 25.1). Transfer abroad is only officially prohibited in cases where advance notification is not given or the transfer results in the loss of a Vietnamese citizen's data. (Article 25.8). Montana's rules lack cross-border and penalty elements, likely in part because American constitutional rules typically forbid states from regulating commerce of an extraterritorial nature, though it does vest "exclusive authority" to enforce the rules in the Montana Attorney General, meaning violations of the rules could result in financial penalties subject to a lawsuit brought on behalf of the state. (Section 12(1)).

In sum, the GMTV cases illustrate rules which are textually and legally different from each other across a range of elements important to rules in the data protection arena. They evince different goals, different notions of the role of the state, textually distinct pre-processing requirements, and different principles as well as unique cross-border rules and penalty schemes. Fundamentally, the rules are distinct in approach. Ultimately, the question is why countries which share so many features would have rules which are legally distinct even as they all share some key terms. Section 6.3 below proposes that domestic law may provide an explanation.

§6.3 Domestic Law as Path Dependence

The meaningful textual and legal differences in data protection canvassed in Section 6.2 correspond to meaningful legal differences at the system level. They thus demonstrate the plausibility of legal distance as a possible factor inhibiting the formation of larger n groups. These legal distinctions include differences in the structure of the courts and sources of law.

Beginning with Gabon, its legal system has inherited some elements of French law as well as the use of French as a legal language.[55] Such elements include specialized Supreme Courts of which it has

[55] The background information presented is derived primarily from the NYU GlobalLex site, which contains country summaries written by domestic lawyers.

six, including two, the High Court of Justice and the Court of Justice of the Republic, dedicated to assessing whether the President or other members of Government respectively commit treason. Like many civil law systems, it also has a specialized Court of Cassation which serves as the court of last resort for civil, criminal, and social matters. A specialized Constitutional Court handles issues of fundamental rights, electoral disputes, conflicts between the various high courts, and the constitutionality of legislation. Outside the realm of family law and succession, which are governed by Gabonese customary law, Gabon tends to adopt or keep legal rules developed elsewhere (see (Jean-Baptiste, 2008)). For illustration, Gabon's Civil Code is based primarily on the French 1804 Code as it existed at Gabonese independence in 1960. And many of its business rules are adapted from the rules developed in conjunction with a regional corporate law harmonizing body, the *Organisation pour l'harmonisation en Afrique du droit des affaires* ("OHADA").

Montana's legal system is based on common law principles and it does not incorporate customary law, unlike Gabon or Tanzania. There is only one court of last resort, the Montana Supreme Court, and as Montana cannot, as a matter of the American Constitution, sign treaties with foreign nations, international harmonization projects have had almost no direct effect on the substance of the law in Montana. Like Montana, Tanzanian law incorporates elements of English common law. Zanzibar remains semi-self-governing except in matters pertaining to its union with mainland Tanzania (formerly Tanganyika). For example, it has its own constitution. But the shared common law heritage of Zanzibar and Tanzania is reflected in their mutual recognition of the received laws of England via the 2002 Judicature and Application of Laws Act and the use of English (supplemented by Kiswahili) as a legal language. Within Zanzibar and certain mainland communities where Islamic norms and customs prevail, common law principles are supplemented with Islamic and customary legal principles. This is particularly true in family law matters. (Rwezaura, 1994, p. 524). Thus, Tanzania is typically viewed as having a "pluralistic" legal system. (Hodgson, 2017, p. 8). This legal pluralism is reflected in its intricate court system with a singular Court of Appeal, functionally the equivalent of a common law-style Supreme Court but without the power of judicial review, two lower courts of unlimited original jurisdiction over cases filed in mainland Tanzania and Zanzibar respectively, a second set of lower appeals courts with appellate jurisdiction only, and finally a series of primary courts at the local level, which in Zanzibar are known as Kadhi Courts and apply Islamic law, except when they sit in matters of criminal law.

Legally, while Viet Nam does have an official civil code dating to the *doi moi* period of liberalization in the late 1980s and early 1990s and a modern commercial code enacted in 2005, as well as an official constitution which applies in its court system, the Vietnamese legal system is not a popular

dispute resolution choice, being relatively slow and not the final legal authority in most cases compared to the CPV. (Greenleaf, 2014, p. 364). Instead, informal private negotiation of disputes (or arbitration for commercial disputes) is the norm, resulting in uneven legal compliance across groups, especially in the growing private commercial sector, (Gillespie, 2015, pp. 962–963). And unlike high courts in Gabon and Montana, the court of last resort in Viet Nam, the People's Supreme Court, does not have the power of judicial review.[56]

These collective legal distinctions lend credence to the supposition that differences across the spectrum of elements of a data protection regime are the result of legal differences between the states which limit the importation of other incompatible models. Systematically stated, the theory that groupings are indicative of path dependence because of shared domestic legal features depends on the assumption that rule choosers are looking for a rule that is best suited to their local characteristics. In a situation where no existing regional model exists, rule choosers must innovate. But where one already exists, that model is typically good enough for the local legal characteristics since those legal characteristics are similar enough to those that exist in the other regional jurisdictions. This would be a form of Herbert Simon-esque "satisficing." (Artinger et al., 2022). And to the extent cross-border regional entities already rely on the existing rule, it would make that choice politically less costly. (Bradford, 2020). Thus, even if the grouped model is not the first choice of the second-adopting jurisdictions or best adapted to its domestic circumstances, it is the choice that has the greatest network effect even over strictly EU models, creating a path dependency where a choice in the first-adopting jurisdiction defines the choice architecture for later-adopting jurisdictions. This could result in a regional grouping without later jurisdictions having necessarily learned that the regional model is optimal.

That story is theoretically defensible. In some ways, the point that law creates textual differences is literally but trivially true. For illustration, Tanzania permits PDPC actions to be appealed to the High Court, the intermediate trial courts for Tanzania and Zanzibar (but presumably only regarding Union matters). (Part VII Subsection 49). That textually distinguishes it from Viet Nam, which makes no mention at all of appeals to a court system or to any court at all, and does not mention a "High Court" because that is not what its highest court is called. Instead, the textual focus is the government ministry that will be making all the decisions as to enforcement, the Ministry of Public Security, and to a lesser extent the

[56] The jurisdictions also diverge on the privacy rights present in their constitutions. This too may drive difference among the textual groupings they belong too. For example, Tanzania protects the right of each person to "the privacy of his own person, his family and of his matrimonial life, and respect" for "his residence and private communications." (Section 16.1). Under the Vietnamese constitution, every person is "entitled to the inviolability of personal privacy, personal secrecy and familial secrecy; and has the right to protect his or her honor and prestige." (Article 21).

Ministry of Information and Communication, which has some regulatory mandate for "media agencies" and the "press" under Article 33.

But in a broader sense, law plausible affects the choice architecture for rule choosers by limiting the "translatability" of certain models developed elsewhere. In part, translatability appears to be proportional to the degree to which the legal limits of each state's territorial power extend. For example, the scope of application appears to be constrained by the limits of state power imposed by law, whether constitutional in the case of Montana, or international in the case of the other three. Thus, Article 4 in Gabon applies the Gabonese rules in any place where Gabonese law applies as a matter of international law. Tanzania uses the similar language. Translatability may also be proportional to overlap in preexisting privacy commitments. To return to a non-GMTV example, Zambia and Zimbabwe share a constitutional commitment to incorporating customary law in their domestic legal regimes (like Gabon and Tanzania) while neither St. Kitts nor Antigua and Barbuda or the U.K. incorporate customary law. And likewise, Zimbabwe in Article 57 has a constitutional commitment to "the privacy of [persons'] communications," which is mirrored by the Zambian constitution in Article 20, Subsection 3b, while neither St. Kitts nor Antigua and Barbuda have comparable rights protections enshrined in their governing constitutions and the U.K. lacks a written constitution.[57] Such legal commitments may have constrained the choice of rule choosers in Zimbabwe to models developed in a country that also has them, in other words, constrained its choice to follow Zambia over St. Kitts or Antigua or the U.K., both for ease of transplantation and to reduce conflict with existing domestic law. Quite clearly domestic law also makes some data protection choices impractical. Consider the fact that some jurisdictions' constitutional and legal arrangements permit independent agencies (e.g., independent DPAs) and others do not. (Greenleaf, 2012).

Furthermore, in any jurisdiction the meaning of texts is partially constructed by the remainder of the legal corpus, adding another layer of choice restraint. This is especially true in the civil law context but it is true in the common law context as well. Consider the *in pari materia* canon of statutory interpretation. In many common law nations, where two statutes "address[] the same subject matter generally they should be read as if they were one law" in order to illuminate a statutory text that is unclear standing alone. (See, e.g., *Wachovia Bank, N.A. v. Schmidt*, 2006).

In short, if such preexisting legal rules shape the translatability of models, data protection could have a path dependent aspect. Once a data protection regime fits well-enough within a shared set of preexisting rules, that is the regime that will tend to be selected in later-adopting jurisdictions whether it

[57] One small exception is that Antigua and Barbuda, in Chapter II, Section 5(3), does protect the privacy of communications between a person who has been arrested or detained and "a legal practitioner of his choice."

is optimal or not. The existence of path dependencies would cast further doubt on the existence of policy learning; the incentive to invest time and resources in finding the best model is decreased commensurate to the ex-ante likelihood that it may conflict with domestic norms and rules, rendering the search costs wasted.

To view the fact that grouping occurs as a reason to question whether rule monoculture exists would consequently be a mistake. Actors could simply be constrained by factors outside their control that hinder their ability to choose a convergent model. Grouping dynamics would instead reveal regional, smaller, and path dependent monocultures that were not the result of the practical observation of outcomes and the selection of the best model. To put it in terms of the biological analogy, rather than one big monoculture of a single species of data protection crop, it seems there are several, disconnected plots. But in each case the plot is dangerously dependent on the viability of a single species.[58]

[58] If that story is true, it would also be a meaningful addition to the comparative law and globalization literatures. Unfortunately, contemporary comparative law focuses on sociological understandings of the paths, motivations, and normative effect of transplants—to the detriment of analyses of law as a barrier to transplantation ((e.g., Miller, 2003; Twining, 2005, pp. 205–206). These are valuable contributions. But they ignore the law aspect of comparative law, namely the limits on convergence through transplantation created by legal rules themselves.

Chapter VII

Rule Diversity Restoration and Research Expansion

It is remarkable that so many jurisdictions have adopted data protection rules; up to 90% of the 193 UN member states have a data protection regime. No doubt, it is tempting to see widespread adoption as a victory for privacy and informational self-determination. If anything, the victory is pyrrhic. And it may not be a victory. Data protection *law* has diffused; data protection may not have. Moreover, the law that has diffused is calcified; the statistical degree of textual similarity between the rules in different jurisdictions is equivalent to benchmark examples of copying. Convergence thus represents a situation of textual monoculture.

Learning, defined as the time-intensive comparison of outcomes of different rules and selection of the more workable rule, is unlikely to be a plausible mechanism behind convergence. Nor is the presence of grouping representative of learning. It instead may indicate domestic-law-based path dependence. For several reasons, including the fact that such homogeneity precludes the possibility of future learning and exists around an EU model which is imperfect, the textual monoculture, and more broadly the rule monoculture it is symptomatic of, is suboptimal.

All is not lost, however. Strategies for preserving and fostering rule diversity exist. They include adopting agnostic data protection rules in trade agreements and international frameworks, and devolving rule making authority in the data protection space. Section 7.2 addresses each of these possibilities in turn, providing evidence that the presence of these features reduces the likelihood of textual homogeneity. The chapter closes, in Section 7.3, with unanswered questions future research could and should pursue. But, before turning to future research, the chapter addresses the conclusions presented over the course of Chapters 1 through 6 in Section 7.1.

§7.1 50 Years of Data Protection Revisited

Global data protection laws have increasingly converged, but that trend is marked by more than a uniform march toward a single set of shared rules. Since the 1970s, statistical distance between first-passed texts—which the German case showed are a valuable proxy for data protection regimes—have declined as measured by Cosine and Jaccard Distances.[59] And the number of distances indicative of

[59] In the same breadth, it is important to explicitly note the dissertation did not directly probe legal meaning, an important part of a rule monoculture. Instead, it focused on legal texts without the type of domestic legal training and acculturation necessary to parse legal meaning. The dissertation is therefore open to Rodolfo Sacco (1991), Günter Frankenberg (2006), and others'

copying and influence compared to benchmark distances between texts with known copying or influence relationships to each other has increased (see Table 3.1 for benchmark values). Visually that convergence appears as a "red shift" towards higher degrees of statistical similarity. The case studies in Chapter 3 illustrate that the "red shift" in textual similarity corresponds to shared text that is legally relevant and meaningful.

In many ways, these are unsurprising results. CDP scholars, even those who conduct little comparative empirical inquiry, are in broad agreement that most jurisdictions' data protection regimes have become more and more similar over time since the 1970s. As Lynskey put it in 2015, "alternatives to the EU's comprehensive regime are fast disappearing at the international level." (Lynskey, 2015, p. 41). But traditional convergence narratives are misleading in some ways and wrong in others. They are misleading in that they fail to address grouping. While the increase in similarity has been modest measured by year-to-year average textual distances, some jurisdictions appear to be growing more similar to each other at a faster rate than jurisdictions overall are becoming more similar to each other. In other words, though it may be true that overall, the distance between the least similar rules continues to decline, albeit slowly, the more salient dynamic since the 2010s is that the textual distances between groups of related jurisdictions is declining. In part, this may be due to the continued outlier status of rules in major markets. For instance, China, California, and India have developed distinct approaches, remaining statistical textual outliers rather than aligning directly with the EU model or any other global standard. And in part this may be due to micro-convergences. For example, countries in East Africa and Spanish-speaking nations in Latin America exhibit notable textual alignment with each other and do not precisely replicate European or American models. These dynamics are reflected visually in the more recent visual trend of darker blocks of red against lighter yellow background dyads.

Traditional convergence narratives are also flatly wrong in some respects. For example, measured by textual similarity it is not correct to argue that there is a broad "D.C. Effect," (Rustad & Koenig, 2019), or a California Effect (Chander et al., 2021), or even to some degree a "Brussels Effect," (Bradford, 2020). Neither broadly American rules nor rules developed in California seem to be widely copied. Instead, Colorado appears to be the originator with other states updating the Colorado template in ways that simplify rules. And GDPR is among the least similar textually to contemporaneous rules. Unlike the DPD, there are no dyads indicating red-range, low Jaccard Distances between GDPR and legislative texts in other jurisdictions. If countries are copying European language, they appear to be

criticism of comparative law as superficial and ignorant of culturally-embedded legal connotations. What the dissertation lacks in depth, it gained in breadth, ultimately revealing patterns in text across the global range of laws.

copying a simplified, significantly pared down version of the legislation, rather than importing European legislation verbatim. Nor does it appear to be the case that the EU data protection model is "easy" to copy simply because it is omnibus in nature. (Schwartz, 2019, p. 818).

The "Queen Model" may encapsulate some of these patterns. Initially, jurisdictions developed unique data protection laws, with limited similarity to each other. France and Luxembourg provide a case in point. Both share critical passages in ways that appear unlikely to be due to random chance in phrasing or word usage. And yet, the two 1970s-vintage legislative texts remain distinctly rooted in national legal predilections. Over time, as the global economy became increasingly interconnected, jurisdictions began adopting elements from each other's frameworks. However, after such harmonizing phases, divergences re-emerged, whether through random chance or domestic innovation or some other factor. Thus, California, China, and India have chosen to take their own idiosyncratic paths. Furthermore, the most recent harmonizing juncture, in theory the GDPR, has not produced the same kind of reconvergence that its predecessor, the DPD, produced. Countries seem to be drawing inspiration from the EU model, but they do not replicate it wholesale as the GMTV and Southern Africa cases indicate. In truth, the GDPR is one of the least textually similar regulations compared to its global contemporaries.

Those results leave the question why jurisdictions are converging and through what mechanism, as well as why groups do not coalesce into larger groups, unanswered. Without resolving the tripartite realism-idealism-incrementalism debate, the data reveals several conclusions. First, the data protection convergence is characterized by a degree of imitation that is not typically present where laws diffuse through learning processes, such as bills of rights, indigenous gaming compacts in American states, and patient classification rules in OECD countries. The fact that the structures through which learning might have occurred are not small world, are siloed and are reliant on EU-based ideas further undermines the inference that learning was a predominant mechanism of convergence. Second, the presence of meaningful legal divergence between the GMTV cases despite their bilateral similarities is evidence that the factors which they all share (e.g., relatively low internet penetration) or share in part (e.g., English as a legal language) are neither necessary nor sufficient for groups to remain textually and legally different from one another. The GMTV cases instead demonstrate the plausibility of legal distance, for example differences in the structure of the courts or sources of law, as a possible factor inhibiting the formation of larger groups. Grouping amongst jurisdictions that share legal similarities may, consequently, be path dependent. Put differently, rule choosers in such jurisdictions may be limited in the degree to which they can transplant alternative models because of preexisting legal commitments.

In sum, statistical textual convergence, in combination with the implausibility of policy learning as a driving mechanism and the possible path dependent nature of rule selection, creates a vicious cycle of diffusion. Learning did not lead to convergence and convergence in turn precludes the possibility of future learning, leading to yet more convergence. And while it may be inevitable that some small degree of diversity in text continues to exist, that small degree of diversity may be due to path dependent dynamics. As such, for jurisdictions that are legally distant and could not select the given rule in any case, the remaining rule diversity is not a precondition for learning. Rule monoculture, of which textual monoculture is a prime warning symptom, consequently has a self-perpetuating effect.

The self-perpetuating nature of the data protection rule monoculture is especially pernicious because there may not be a "right" way to do data protection. Privacy norms vary around the world and so do the costs and benefits of information rights. Even if there is a single "right" way to do data protection, the EU model may not be it. Its mandate to facilitate data trade as well as its intentional vagueness may render it ill-suited to address the massive conglomeration of economic power wrought by data.

§7.2 Preserving Rule Diversity

The obvious follow-on question is how to preserve what rule diversity exists and grow it going forward. From the data collected here, there appear to be two options. First, jurisdictions could adhere to agnostic cross-border data trade rules, such as those found in the United-States-Mexico-Canada Agreement ("USMCA") and the APEC Privacy Framework. A second, and perhaps more promising, option is to devolve rule-making authority to the constituent parts of states.[60] Each option comes with tradeoffs but appears to have the effect of sheltering rule diversity.

The USMCA provisions and the APEC Framework do not impose any particular data protection regime on their signatories. Article 19.8.2 of the USMCA commits the parties to maintaining "a legal framework that provides for the protection of the personal information of the users of digital trade." It requires the parties to take account of OECD principles and the APEC Privacy Framework but it stops short of requiring a specific interpretation or set of rules, instead listing exemplar principles to consider like "use limitation," "purpose specification," "transparency," and "accountability." (Article 19.8.3). This detachment from specific rules is part and parcel of Article 19.8's general goal, which is to ensure that

[60] A third option that might spring to mind is to reinterpret the EU law of adequacy. But that might not be as promising for a couple of reasons. First, most adopting countries have no chance of receiving an adequacy decision regardless of what rules they have in place because their data is not valuable to the EU for economic or geo-political reasons. Adjusting adequacy would therefore have little theoretical effect on such countries. Second, there are reasons to believe that in practice "essentially equivalent" remains a flexible standard. (Schwartz, 2019).

"any restrictions on cross-border flows of personal information are necessary and proportionate to the risks presented." (Article 19.8.3). Thus, "each party," is encouraged to "promote compatibility between" their regimes. (Article 19.8.6).

The APEC cross border rules ("CBPR") similarly prioritize compatibility while avoiding imposing a specific set of rules on member states. CBPR requires only that information controllers in contracting jurisdictions be certified as compliant by an independent "Accountability Agent" backed up, in theory, by a local enforcement agency which handles consumer complaints. (Sullivan, 2019, pp. 5–7). The APEC framework parallels the USMCA in noting that contracting parties "should refrain from restricting cross border flows of personal information … where (a) the other economy has in place legislative or regulatory instruments that give effect to the Framework or (b) sufficient safeguards exist," such as compliance with the CBPR scheme. (APEC Framework, 2015, ¶ 69). The obvious downside of such agnostic rules is that trade is prioritized over the protection of personal information. (Chander & Schwartz, 2023, p. 6).[61]

But the USMCA and APEC Framework are good for rule diversity. In fact, they shelter it. To be sure, the direction of causality is unclear. Still, regardless of whether divergent rules necessitated rule-agnostic trade agreements or such agreements allowed for divergent rules, it cannot be said that the existing trade agreements have quashed rule diversity. That is a rarity in the data protection space. Canada, Mexico, and the existing American state rules are all very different from one another when it comes to currently in force data protection rules. As a statistical matter, American rules are very dissimilar from Mexican and Canadian rules (in general, Cosine > .9 and Jaccard > .7) and the Mexican and Canadian rules, while less distinct (e.g., Jaccard >.6), remain different from each other. Such distance values approach the distance values of rules picked at random from the same area of law (see Table 3.1).

To give just a few qualitative examples of divergence, Mexico's cross-border rules require only provision of a privacy notice and the purpose behind collection to the transferee as well as the consent of the data subject (Article 36) while Canada's rules make the transferor responsible for ensuring the transferee organization (rather than state) has a "comparable level of protection," (Schedule 1, Section 4.1.3). Assuming that the purpose of processing is the same in the transferee organization, additional consent need not be secured, in contrast to Mexico. For their part, the American state-level rules do not have cross-border provisions. Another example is each jurisdiction's approach to defining consent to

[61] Chander and Schwartz call for a Global Agreement on Privacy that would establish baseline rules that are shared in as many countries as possible, thereby facilitating trade while avoiding the unnecessarily lax USMCA and APEC approach. (Chander & Schwartz, 2023, p. 47). But that would have the obvious downside of completely and utterly ensuring rule monoculture.

information processing. The Canadian rules do not define consent and instead give examples of possible routes for obtaining it (e.g., Schedule 1, Section 4.3.6 and 4.3.7) and situations in which consent cannot be obtained (e.g., Schedule 1, Section 4.3.3). The Mexican rules also eschew a specific definition of consent, opting to define it through examples and, in contrast to the Canadian approach, to permit "tacit" consent. (Article 7). The American state-level rules invariably define consent specifically. Illustratively, Iowa defines consent as "a clear affirmative act signifying a consumer's freely given, specific, informed, and unambiguous agreement." (Section 1.6). In each example, the jurisdictions diverge, despite being geographically close to one another, which is unusual compared to the results in Chapter 3.

A second option is devolution, which also shelters rule diversity. Devolution might work best when centralization has not yet occurred. In such situations, devolution avoids the homogeneity that centralization can create by harmonizing rules across jurisdictions subject to the authority of the central rule-maker. Thus, the rules at the American state level are distinct from one another and especially from California's rules. One can see this most clearly in Case Study 2 in Part II of Chapter 3 comparing legally relevant text across American states. But it is also apparent on heatmaps, particularly Figures 3.13, 3.14, and 3.15. In each time bracket, patterns across American states indicate a degree of continued rule diversity, even as some textual distances between states appear to be consonant with influence compared to benchmarks (see Table 3.1). One reason why devolution may shelter rule diversity is that privacy cultures do tend to differ across communities. And though that may not be true to a meaningful extent across American states, it certainly is true across member states of the EU, (Whitman, 2004), and across jurisdictions globally, (Arora, 2019).

Of course, devolution of rule-making authority is controversial in the privacy literature. Most scholars are against it. Petkova, for example, argues that federalism can in some cases promote pro-privacy rules but concedes that privacy federalism is not inherently good. Thus, centralization often cuts through a patchwork "privacy thicket" by solidifying the gains of competitive rulemaking dynamics between vertical government structures. (2016, p. 600). She adds that the "centralization of privacy [rules] does not carry with it the risk of ossification," provided that the rules remain democratically accountable. (Petkova, 2016, p. 600). Schwartz concurs. In his view, there is "too little consolidation" in the American context because of the lack of action at the federal level, leading to a confusing and counterproductive mix of state-level regulation. (Schwartz, 2015, p. 343). Goldman adds that devolution in the American context is a bad idea because state legislatures are easier to capture than Congress and are often less well-equipped to write decent rules, there are no "local" conditions of privacy which justify

devolution, and compliance cost increases outweigh any social value from experimentation. (Goldman, 2015).[62]

Such concerns are likely overblown. Centralization does not always assist in trimming the privacy thicket. The quintessential example of centralization—the EU—retains its own "thicket" of rules regulating the information ecosystem including the 2002 ePrivacy Directive [2002/58/EC], the AI Act [Regulation 2024/1689], and GDPR. Nor is it necessarily true, as Goldman argues, that legislatures lower down the federal scale have less expertise because they are less likely to deal with privacy issues or be subject to privacy lobbying. In fact, bespoke, expert privacy lobbying may have catalyzed the data protection regime in California. (Chander et al., 2021, p. 1782). And the absence of democratic accountability is no less a problem at higher levels within federal states, contrary to Petkova's position. (Smith, 2007, p. 39). At heart, the gains in rule diversity may be worth the marginal risks identified by scholars who disfavor privacy federalism. Certainly, as Section 7.3 suggests, there is much analysis yet to be done.

§7.3 Future Research Directions

The findings in Section 7.1 are cabined to data protection law. Simply put, data protection law has some unique aspects that other areas of law do not: These include an inherently transnational nature, a fast-evolving regulatory environment, and relatively recent provenance. What is more, the data collected related only to data protection law and so any conclusions about legal diffusion writ large or in other areas of law would be speculative.

That highlights the necessity of expanding the research in the future. Expansion could take one of three tracks: expand the methods, expand the data, or expand the scope of laws covered. Methodologically, there are several possible research extensions. Additional similarity metrics could be employed (e.g., Tversky Index). AI-enabled word embedding techniques could be employed to get at semantic rather than strictly textual similarity, albeit at the cost of additional required computing power. Alternatively, Cosine and Jaccard Distances could be computed for each rule and every rule after it iteratively over the entire rule set. Although it would be time and computation power intensive, that approach would better reveal the provenance of data protection ideas than time-bracketing. Along similar lines, the network analysis could be deepened by additional community detection techniques, which could reveal granular clustering phenomena in the graphs.

[62] Of course, there are also practical concerns with devolution such as the fact that an EU-level rule may be the only feasible European policy choice given member states' need to balance security and market growth, (Mendez & Mendez, 2010, p. 640). The implication of Mendez and Mendez's argument is that the feasible political options in Europe are EU-level rules or, in many states, no rules at all.

Additional qualitative approaches could also be employed, including interviews with rule choosers and/or alternative conceptualizations of learning. Interviews might help establish the conscious reasons rule choosers had for adopting a data protection rule in a given case, while one alternative way to measure learning might be the non-adoption of rules that are known to have "failed" in practice. Using those methods would of course entail new challenges. What it means for a rule to "fail" and how failure could be observed are, for instance, complex questions. And neither interview candor nor access to rule choosers is guaranteed.

A further approach would be to add data. All rules, rather than just the first-passed rule could be collected. Or online traces of connections between regulators could be scraped to increase the size of the social network graph presented in Chapter 5 Part II. And the size and range of the benchmark table in Section 3.3 could be increased to provide a more thorough view of what influence and copying look like in textual distance. Moreover, data protection is not the only type of rule that affects personal privacy. And so, the dissertation could be expanded to address, for example, data breach rules, communications privacy rules, constitutional privacy provisions, or surveillance rules, without departing from the underlying theme of exploring patterns in global privacy law. Doing so might paint a more vibrant picture of the global diffusion of privacy ideas than a focus on data protection alone.

Regardless of possible extensions, the findings presented have important implications for the study of legal diffusion. For one thing, they highlight the returns to pursuing empirical approaches. Not only does the empirical approach help illuminate the what and where of convergence, it can also help illuminate the why of convergence by distinguishing between emulation and learning. This opens a whole panoply of possible research questions. Are legal regimes globally converging? Is copying or learning more common when states adopt new laws? Which jurisdictions are most likely to copy as opposed to learn and vice versa? How common are rule monocultures? What does it take to disrupt them? Such questions strike at the heart of the role of law in a globalized world. While global patterns in data protection law do not reveal answers to such questions, they provide an excellent reason to start looking.

Appendix A – Selected Matrix Values

Countries (row, column)	Value
1 Barbados, Belize	.357
2 St. Kitts, Antigua and Barbuda	.368
3 Virginia, Iowa	.465
4 Virginia, Connecticut	.473
5 Colorado, Connecticut	.546
6 Connecticut, Iowa	.553
7 Israel, Bahamas	.558
8 Kenya, Mauritius	.565
9 Madagascar, Senegal	.566
10 Senegal, South Africa	.570

Table Appendix A.1. Jaccard Distances of the top 10 most similar rules by Jaccard Distance.

Countries (symmetrical)	Value
1 Belize, Barbados	0.01833686
2 Iowa, Virginia	0.03982198
3 Antigua and Barbuda, St. Kitts	0.04188001
4 Connecticut, Virginia	0.04388673
5 Kenya, Zambia	0.04805721
6 Botswana, Barbados	0.05837916
7 Connecticut, Colorado	0.05993431
8 Iowa, Connecticut	0.06358818
9 Colorado, Virginia	0.06420551
10 Kenya, Botswana	0.06806361

Table Appendix A.2. Cosine Distances of the top 10 most similar rules by Cosine Distance.

Full matrices are available as .csv files from the author. They are unlabeled but a key is available for download from https://github.com/ccrum1/dissertation.

Appendix B – Selected Additional Figures

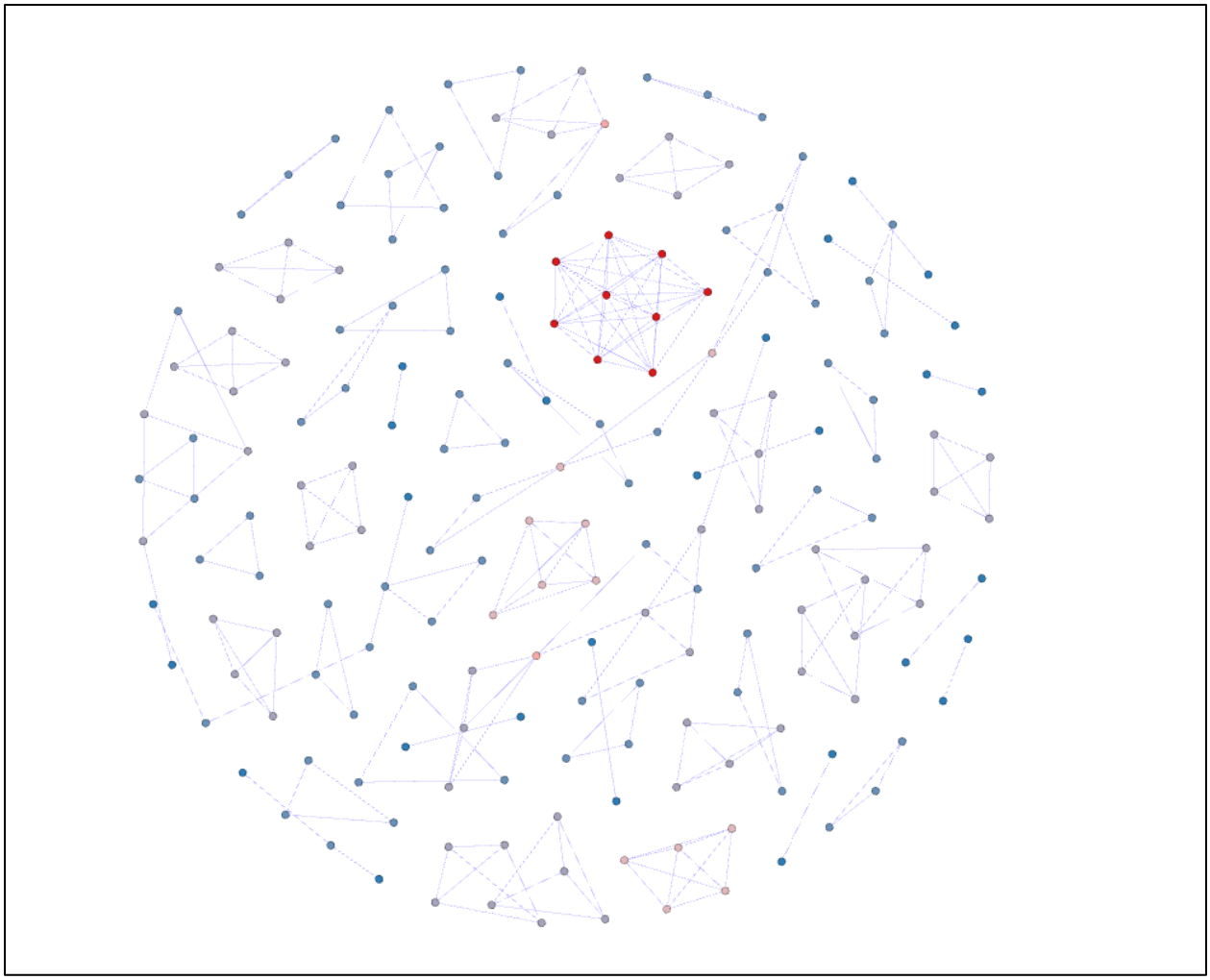


Figure B.1. Network graph of attendees at the 2024 Privacy and Security Forum. Color of node indicates degree (gray = low degree, blue = average degree, red = high degree).

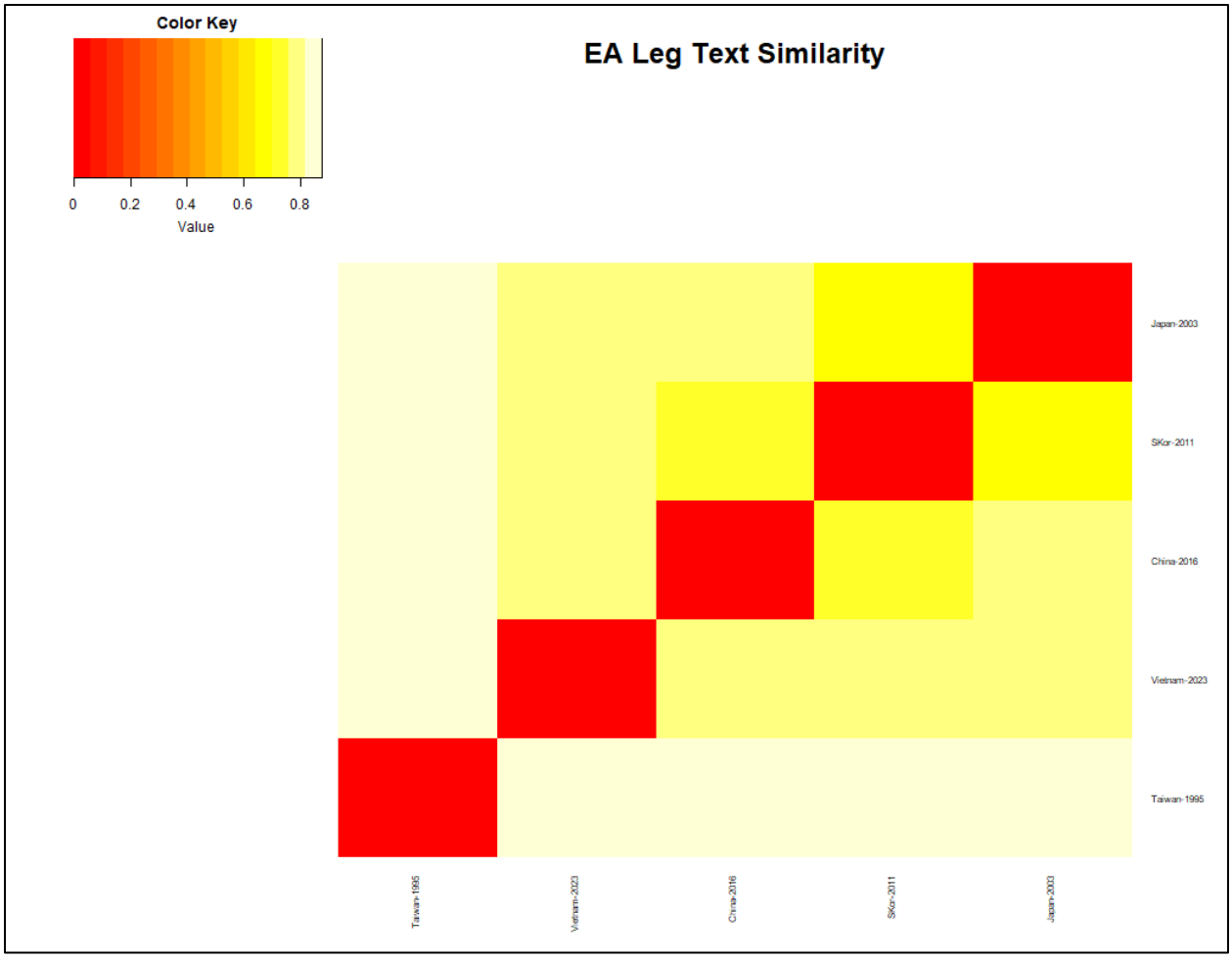


Figure B.2. Heatmap of East Asian data protection texts by Jaccard Distance. Colors toward the red end of the spectrum indicate higher degrees of similarity.

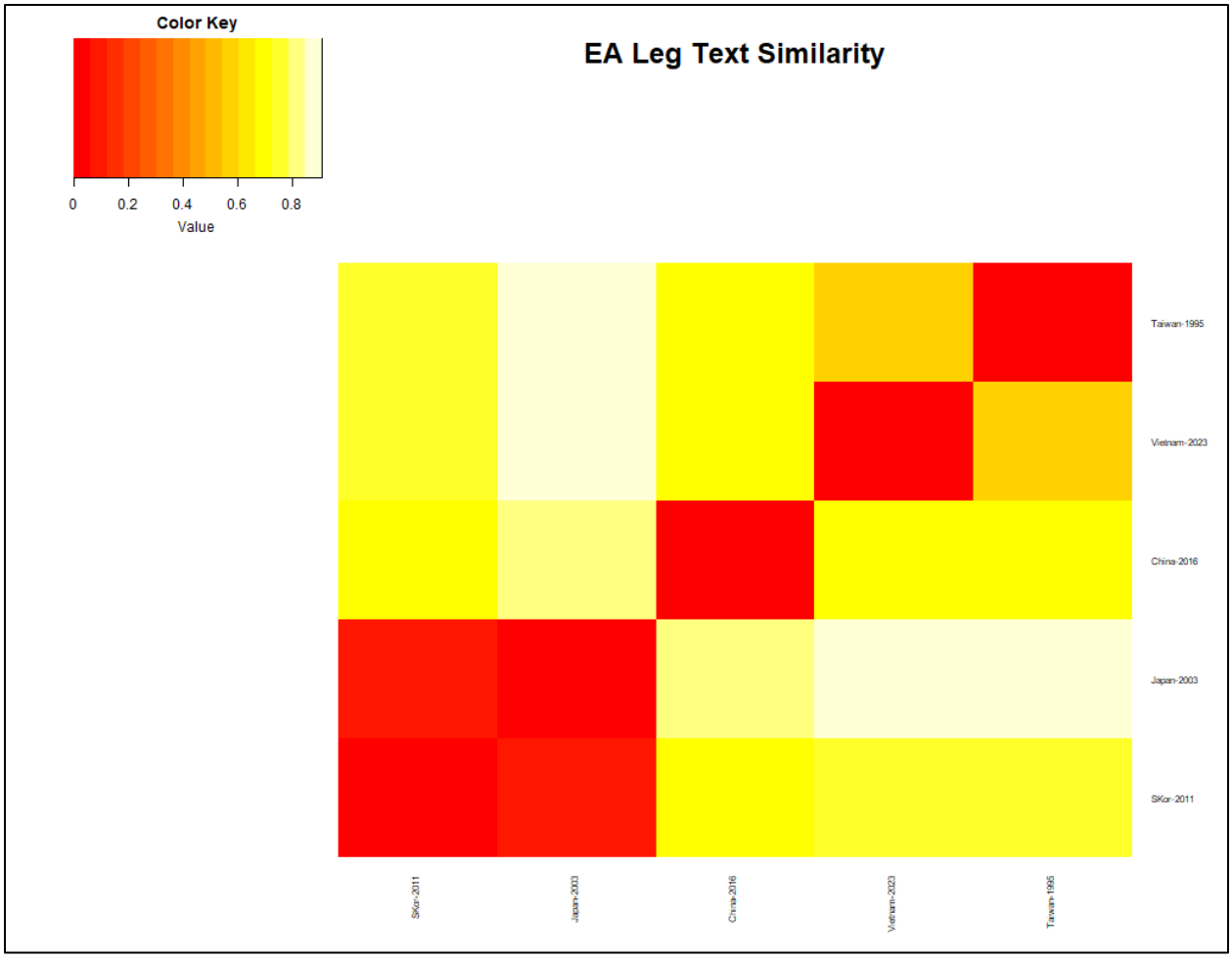


Figure B.3. Heatmap of East Asian data protection texts by cosine distance. Colors toward the red end of the spectrum indicate higher degrees of similarity.

Bibliography

Acquisti, A., John, L. K., & Loewenstein, G. (2013). What is Privacy Worth? *The Journal of Legal Studies*, *42*(2), 249–274.

Acquisti, A., Taylor, C., & Wagman, L. (2016). The Economics of Privacy. *Journal of Economic Literature*, *54*(2), 442–492.

Akee, R. K. Q., Spilde, K. A., & Taylor, J. B. (2015). The Indian Gaming Regulatory Act and Its Effects on American Indian Economic Development. *Journal of Economic Perspectives*, *29*(3), 185–208. https://doi.org/10.1257/jep.29.3.185

Alschner, W. (2020). Sense and Similarity: Automating Legal Text Comparison. In R. Whalen (Ed.), *Computational Legal Studies: The Promise and Challenge of Data-driven Research*. Edward Elgar.

Alschner, W., Panford-Walsh, R., & Skougarevskiy, D. (2018). What Can the Negotiations of NAFTA 1.0 Teach Us About the Fate of NAFTA 2.0? *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3123427

Alschner, W., & Skougarevskiy, D. (2016). Mapping the Universe of International Investment Agreements. *Journal of International Economic Law*, *19*(3), 561–588. https://doi.org/10.1093/jiel/jgw056

Amaral, L. A. N., Scala, A., Barthélémy, M., & Stanley, H. E. (2000). Classes of small-world networks. *Proceedings of the National Academy of Sciences*, *97*(21), 11149–11152. https://doi.org/10.1073/pnas.200327197

Andrea, J., & Wiewiórowski, W. (2022, September 12). *Open Letter on EDPB Budget Proposal for 2023*. https://edpb.europa.eu/system/files/2022-09/letteronbudget_out2022-0068.pdf

Arora, P. (2019). Decolonizing Privacy Studies. *Television & New Media*, *20*(4), 366–378. https://doi.org/10.1177/1527476418806092

Artinger, F. M., Gigerenzer, G., & Jacobs, P. (2022). Satisficing: Integrating Two Traditions. *Journal of Economic Literature*, *60*(2), 598–635. https://doi.org/10.1257/jel.20201396

Bamberger, K. A., & Mulligan, D. K. (2015). *Privacy on the Ground: Driving Corporate Behavior in the United States and Europe*. MIT Press.

Barabási, A.-L., & Albert, R. (1999). Emergence of Scaling in Random Networks. *Science*, *286*(5439), 509–512. https://doi.org/10.1126/science.286.5439.509

Barak-Erez, D. (2009). The Institutional Aspects of Comparative Law. *Columbia Journal of European Law*, *15*(3), 477–494.

Bedner, A. (2013). Indonesian Legal Scholarship and Jurisprudence as an Obstacle for Transplanting Legal Institutions. *Hague Journal on the Rule of Law*, *5*(2), 253–273.

Belli, L., & Doneda, D. (2022). Data protection in the BRICS countries: Legal interoperability through innovative practices and convergence. *International Data Privacy Law*, 1–23. https://doi.org/10.1093/idpl/ipac019

Bennett, C. J. (1992). *Regulating Privacy—Data Protection and Public Policy in Europe and the United States*. Cornell University Press.

Bennett, C. J. (1997a). Convergence Revisited: Toward a Global Policy for the Protection of Personal Data? In P. E. Agre & M. Rotenberg (Eds.), *Technology and Privacy: The New Landscape* (pp. 99–103). MIT Press.

Bennett, C. J. (1997b). Understanding Ripple Effects: The Cross–National Adoption of Policy Instruments for Bureaucratic Accountability. *Governance*, *10*(3), 213–233. https://doi.org/10.1111/0952-1895.401997040

Bennett, C. J. (2008). *The Privacy Advocates: Resisting the Spread of Surveillance*. MIT Press.

Bennett, C. J. (2010). *The Privacy Adovcates: Resisting the Spread of Surveillance*. MIT Press.

Bennett, C. J. (2018). The European General Data Protection Regulation: An instrument for the globalization of privacy standards? *Information Polity*, *23*(2), 239–246. https://doi.org/10.3233/IP-180002

Bennett, C. J., & Howlett, M. (1992). The Lessons of Learning: Reconciling Theories of Policy Learning and Policy Change. *Policy Sciences*, *25*(1), 275–294.

Bennett, C. J., & Raab, C. (2006). *The Governance of Privacy: Policy Instruments in Global Perspective*. MIT Press.

Bennett, C. J., & Raab, C. D. (2020). Revisiting the governance of privacy: Contemporary policy instruments in global perspective. *Regulation & Governance*, *14*(3), 447–464. https://doi.org/10.1111/rego.12222

Berkowitz, D., Pistor, K., & Richard, J.-F. (2003). The Transplant Effect. *American Journal of Comparative Law*, *51*, 163–204.

Bianchi, A. (2019). Epistemic Communities. In *Concepts for International Law*.

Bing, J. (1978). A comparative outline of privacy legislation. In *Comparative Law Yearbook* (pp. 149–181). Sijthoff & Noordhoff.

Bing, J. (1999). Data Protection, Jurisdiction and the Choice of Law. *Privacy Law and Policy Reporter*, *65*(6(6)), 1–10.

Birman, K. P., & Schneider, F. B. (2009). The Monoculture Risk Put into Context: IEEE Xplore. *IEEE Security & Privacy*, *71*(1), 14–17.

Boehmke, F. J., & Witmer, R. (2004). Disentangling Diffusion: The Effects of Social Learning and Economic Competition on State Policy Innovation and Expansion. *Political Research Quarterly*, *57*(1), 39–51. https://doi.org/10.1177/106591290405700104

Boulet, R., Mazzega, P., & Bourcier, D. (2011). A network approach to the French system of legal codes—part I: Analysis of a dense network. *Artificial Intelligence and Law*, *19*(4), 333–355. https://doi.org/10.1007/s10506-011-9116-1

Boushey, G. (2012). Punctuated Equilibrium Theory and the Diffusion of Innovations. *The Policy Studies Journal*, *40*(1), 127–146.

Bowman, G. (2006). The art of legislative drafting. *Amicus Curiae*, 2–9. https://doi.org/10.14296/ac.v2006i64.1117

Bracha, O. (2010). The Adventures of the Statute of Anne in the Land of Unlimited Possibilities: The Life of a Legal Transplant. *Berkeley Technology Law Journal*, *25*(1), 1427–1473.

Bracy, J. (2024, October 1). 2024 in US state privacy law: A tretrospective with Keir Lamont and David Stauss. *IAPP*. https://iapp.org/news/a/2024-in-us-state-privacy-law-a-retrospective-with-keir-lamont-and-david-stauss

Bradford, A. (2012). The Brussels Effect. *Northwestern University Law Review*, *107*(1), 1–68.

Bradford, A. (2015). Exporting standards: The externalization of the EU's regulatory power via markets. *International Review of Law and Economics*, *42*(1), 158–173.

Bradford, A. (2020). *The Brussels Effect: How the European Union Rules the World*. Oxford University Press.

Bradford, A., Chang, Y., Chilton, A., & Garoupa, N. (2021). Do Legal Origins Predict Legal Substance. *Journal of Law & Economics*, *64*(2), 207–232.

Bradford, A., Chilton, A., & Linos, K. (2023). The Gravity of Legal Diffusion. *University of Chicago Legal Forum*, *2023*, 35–57.

Burgess, M., Giraudy, E., Katz-Samuels, J., Walsh, J., Willis, D., Haynes, L., & Ghani, R. (2016). The Legislative Influence Detector: Finding Text Reuse in State Legislation. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 57–66. https://doi.org/10.1145/2939672.2939697

Burkert, H. (1981). Institutions of Data Protection—An Attempt at a Functional Explanation of European National Data Protection Laws. *Computer Law Journal*, *3*(1), 167–188.

Burri, M. (2023). Cross-border data flows and privacy in global trade law: Has trade trumped data protection? *Oxford Review of Economic Policy*, *39*(1), 85–97. https://doi.org/10.1093/oxrep/grac042

Bygrave, L. A. (2014). *Data Privacy Law: An International Perspective*. Oxford University Press.

Bygrave, L. A., & Yeung, K. (2022). Demystifying the modernized European Data Protection Regime: Cross-Disciplinary Insights from Legal and Regulatory Governance Scholarship. *Regulation and Governance*, *16*(1), 137–155.

Cai, P., & Chen, L. (2022). Demystifying data law in China: A unified regime of tomorrow. *International Data Privacy Law*, *12*(2), 75–92. https://doi.org/10.1093/idpl/ipac004

Cairns, J. W. (2013). Watson, Walton, and the History of Legal Transplants. *Georgia Journal of International and Comparative Law*, *41*(1), 637–696.

Cate, F. H. (1997). *Privacy in the Information Age*. The Brookings Institution.

Chander, A., Kaminski, M. E., & McGeveran, W. (2021). Catalyzing Privacy Law. *Minnesota Law Review*, *105*(1), 1733–1802.

Chander, A., & Lê, U. P. (2015). Data Nationalism. *Emory Law Journal*, *64*(1), 677–739.

Chander, A., & Schwartz, P. M. (2023). Privacy and/or Trade. *University of Chicago Law Review*, *90*(1).

Chané, A.-L., & Sharma, A. (2017). Social network analysis in human rights research. In B. A. Andreassen, H.-O. Sano, & S. McInerney-Lankford (Eds.), *Research Methods in Human Rights*. Edward Elgar Publishing. https://doi.org/10.4337/9781785367793.00023

Christo, E. D. (2013). Data protection in Trinidad and Tobago. *International Data Privacy Law*, *3*(3), 202–209. https://doi.org/10.1093/idpl/ipt013

CISCO. (2021). *Cisco 2021 Consumer Privacy Survey: Building Consumer Confidence Through Transparency and Control*.

Cofone, I. (2021). Beyond Data Ownership. *Cardozo Law Review*, *43*(2), 501–572.

Cofone, I. (2023). *The Privacy Fallacy: Harm and Power in the Information Economy*. Cambridge University Press.

Cohn, M. (2010). Legal Transplant Chronicles: The Evolution of Unreasonableness and Proportionality Review of the Administration in the United Kingdom. *American Journal of Comparative Law*, *58*(3), 583–630.

Coll, S. (2014). Power, knowledge, and the subjects of privacy: Understanding privacy as the ally of surveillance. *Information, Communication, and Society*, *17*(10), 1250–1263.

Council of Europe. (1981, January 28). *Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data*.

Cranor, L. F. (2002). The role of privacy advocates and data protection authorities in the design and deployment of the platform for privacy preferences. *Proceedings of the 12th Annual Conference on Computers, Freedom and Privacy*, 1–8. https://doi.org/10.1145/543482.543506

Curran, V. G. (2019). Comparative Law and Language. In M. Reimann & R. Zimmermann (Eds.), *The Oxford Handbook of Comparative Law* (pp. 680–709). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198810230.013.21

Dallara, C., & Piana, D. (2016). *Networking the Rule of Law* (0 ed.). Routledge. https://doi.org/10.4324/9781315597812

Daniels, R. J., Trebilcock, M. J., & Carson, L. D. (2011). The Legacy of Empire: The Common Law Inheritance and Commitments to Legality in Former British Colonies. *American Journal of Comparative Law*, *591*(1), 111–178.

Daskal, J. (2015). The Un-Territoriality of Data. *Yale Law Journal*, *125*(2), 326–399.

De Almeida Costa, A. B. (2023). Citation networks in administrative law books from the civil law world (nineteenth century). *Comparative Legal History*, *11*(2), 154–185. https://doi.org/10.1080/2049677X.2023.2270388

De Terwangne, C. (2020). Article 5 Principles relating to processing of personal data. In C. Kuner, L. A. Bygrave, C. Docksey, & L. Drechsler (Eds.), *The EU General Data Protection Regulation (GDPR)* (pp. 309–320). Oxford University PressNew York. https://doi.org/10.1093/oso/9780198826491.003.0034

De Visser, M., & Claes, M. (2013). Judicial Networks. In P. Larouche & P. Cserne (Eds.), *National Legal Systems and Globalization* (pp. 345–368). T. M. C. Asser Press. https://doi.org/10.1007/978-90-6704-885-9_13

Deakin, S., Gindis, D., Hodgson, G. M., Kainan, H., & Pistor, K. (2017). Legal Institutionalism: Capitalism and the Constitutive Role of Law. *Journal of Comparative Economics*, *45*(1), 188–200. https://doi.org/10.1016/j.jce.2016.04.005

Del Duca, L. F., & Levasseur, A. A. (2010). Impact of Legal Culture and Legal Transplants on the Evolution of the U.S. Legal System. *American Journal of Comparative Law*, *58*(1), 1–30.

Denny, M. J., & Spirling, A. (2018). Text Preprocessing For Unsupervised Learning: Why It Matters, When It Misleads, And What To Do About It. *Political Analysis*, *26*(2), 168–189. https://doi.org/10.1017/pan.2017.44

Dodds, P. S., Watts, D. J., & Sabel, C. F. (2003). Information exchange and the robustness of organizational networks. *Proceedings of the National Academy of Sciences*, *100*(21), 12516–12521. https://doi.org/10.1073/pnas.1534702100

Dolowitz, D. P. (2000). Special Issue on Policy Transfer—Introduction. *Governance*, *13*(1), 1–4. https://doi.org/10.1111/0952-1895.00120

Dothan, S. (2018). Social networks and the enforcement of international law. In M. Hirsch & A. Lang (Eds.), *Research Handbook on the Sociology of International Law*. Edward Elgar Publishing. https://doi.org/10.4337/9781783474493.00023

Drezner, D. W. (2008). *All Politics is Global: Explaining International Regulatory Regimes*. Princeton University Press.

Dunlop, C. (2013). Policy transfer as learning: Capturing variation in what decision-makers learn from epistemic communities. In M. Evans (Ed.), *New Directions in the Study of Policy Transfer*. Taylor and Francis.

Dunlop, C. A., & Radaelli, C. M. (2013). Systematising Policy Learning: From Monolith to Dimensions. *Political Studies*, *61*(3), 599–619.

Dunlop, C. A., & Radaelli, C. M. (2020). Policy Learning. In *The Sage Handbook of Political Science*. SAGE Publications.

Düpont, N., & Rachuj, M. (2022). The Ties That Bind: Text Similarities and Conditional Diffusion among Parties. *British Journal of Political Science*, *52*(2), 613–630. https://doi.org/10.1017/S0007123420000617

Edelman, L. B. (2016). *Working Law: Courts, Corporations, and Symbolic Civil Rights*. University of Chicago Press.

Engel, C., & Weinshall, K. (2022). Diffusion of Legal Innovations. *Annual Review of Law and Social Science*, *18*(1), 139–153. https://doi.org/10.1146/annurev-lawsocsci-050420-012835

Erdos, D. (2016). European Data Protection Regulation and Online New Media: Mind the Enforcement Gap. *Journal of Law and Society*, *43*(4), 534–564.

Ess, C. (2006). Ethical pluralism and global information ethics. *Ethics and Information Technology*, *8*(4), 215–226.

Evans, M. (2010). Introduction: New directions in the study of policy transfer. In *New Directions in the Study of Policy Transfer*. Routledge.

Farrell, H., & Newman, A. L. (2015). The New Politics of Interdependence: Cross-National Layering in Trans-Atlantic Regulatory Disputes. *Comparative Political Studies*, *48*(4), 497–526. https://doi.org/10.1177/0010414014542330

Farrell, H., & Newman, A. L. (2019). *Of Privacy and Power: The Transatlantic Struggle over Freedom and Security*. Princeton University Press.

Finck, M., & Pallas, F. (2020). They who must not be identified—Distinguishing personal from non-personal data under the GDPR. *International Data Privacy Law*, *10*(1), 11–36. https://doi.org/10.1093/idpl/ipz026

Fiodorova, A. (2018). *Information Exchange and EU Law Enforcement* (1st ed.). Routledge. https://doi.org/10.4324/9781351240260

Fischel, D. (1981). Race to the Bottom Revisited: Reflections on Recent Developments in Delaware's Corporation Law. *Northwestern University Law Review*, *76*(6), 913–945.

Flaherty, D. H. (1979). *Privacy and Government Data Banks: An International Perspective*. Mansell Publishing.

Flaherty, D. H. (1989). *Protecting Privacy in Surveillance Societies: The Federal Republic of Germany, Sweden, France, Canada, and the United States*. University of North Carolina Press.

Flaherty, D. H., Donohue, T. J., & Harte, P. J. (Eds.). (1984). *Privacy and Data Protection: An International Bibliography*.

Fourcade, M., & Gordon, J. (2020). Learning Like a State: Statecraft in the Digital Age. *Journal of Law and Political Economy*, *1*(1), 78–108.

Fowler, J. H., Johnson, T. R., Spriggs, J. F., Jeon, S., & Wahlbeck, P. J. (2007). Network Analysis and the Law: Measuring the Legal Importance of Precedents at the U.S. Supreme Court. *Political Analysis*, *15*(3), 324–346. https://doi.org/10.1093/pan/mpm011

Frankenberg, G. (2006). Comparing constitutions: Ideas, ideals, and ideology—toward a layered narrative. *International Journal of Constitutional Law*, *4*(3), 439–459. https://doi.org/10.1093/icon/mol012

Frankenberg, G. (2010). Constitutional transfer: The IKEA theory revisited. *International Journal of Constitutional Law*, *8*(3), 563–579. https://doi.org/10.1093/icon/moq023

Frankenreiter, J. (2021). The Missing "California Effect" in Data Privacy Law. *Legal Studies Research Paper Series*, *Paper No. 21-07-01*(July 2021), 1–64.

Fried, C. (1968). *Privacy*. *77*(3), 475–493.

Fuster, G. G. (2014). *The Emergence of Personal Data Protection as a Fundamental Right of the EU*. Springer.

Garoupa, N., & Ogus, A. (2006). A Strategic Interpretation of Legal Transplants. *Journal of Legal Studies*, *35*(1), 339–364.

Garrett, K. N., & Jansa, J. M. (2015). Interest Group Influence in Policy Diffusion Networks. *State Politics & Policy Quarterly*, *15*(3), 387–417.

Gavison, R. (1980). Privacy and the Limits of Law. *Yale Law Journal*, *89*(1), 421–471.

Gbadamosi, N. (2023, September 1). Has Gabon's "all-powerful" Bongo dynasty really lost its 55-year grip? *Al Jazeera*.

GDMA. (2022). *Global Data Privacy: What the Consumer Really Thinks*. https://dma.org.uk/research/global-data-privacy-what-the-consumer-really-thinks-1#:~:text=In%202022%2C%20almost%20half%20(46,stated%20this%20back%20in%202018

Gellman, R. (2022, April 6). *Fair Information Practices: A Basic History*. https://bobgellman.com/rg-docs/rg-FIPShistory.pdf

Gilardi, F. (2010). Who Learns from What in Policy Diffusion Processes? *American Journal of Political Science*, *54*(3), 650–666. https://doi.org/10.1111/j.1540-5907.2010.00452.x

Gilardi, F., Füglister, K., & Luyet, S. (2009). Learning From Others: The Diffusion of Hospital Financing Reforms in OECD Countries. *Comparative Political Studies*, *42*(4), 549–573. https://doi.org/10.1177/0010414008327428

Gillespie, J. (2015). Localized Global Competition Law in Vietnam: A Bottom-up Perspective. *International and Comparative Law Quarterly*, *64*(4), 935–964.

Ginsburg, T. (2002). Confucian Constitutionalism—The Emergence of Constitutional Review in Korea and Taiwan. *Law & Social Inquiry*, *27*(4), 763–800.

Goldbach, T. S. (2019). Why Legal Transplants? *Annual Review of Law and Social Science*, *15*(1), 583–601. https://doi.org/10.1146/annurev-lawsocsci-101518-042617

Goldman, E. (2015, September 2). The Long-term Promise of Privacy Federalism, Part 2. *Technology and Marketing Law Blog*. https://blog.ericgoldman.org/archives/2015/09/the-long-term-promise-of-privacy-federalism-part-2.htm

Google Transparency Report. (2023, March 20). *Delisting URLs from Google Search for privacy*. Requests to Delist Content under European Privacy Law. https://transparencyreport.google.com/eu-privacy/overview?hl=en_GB

Graham, E. R., Shipan, C. R., & Volden, C. (2013). The Diffusion of Policy Diffusion Research in Political Science. *British Journal of Political Science*, *43*(3), 673–701. https://doi.org/10.1017/S0007123412000415

Grande, E. (2000). Italian Criminal Justice: Borrowing and Resistance. *American Journal of Comparative Law*, *48*, 227–260.

Grande, E. (2016). Legal Transplants and the Inoculation Effect: How American Criminal Procedure Has Affected Continental Europe. *American Journal of Comparative Law*, *64*(3), 583–618.

Gray, V. (1973). Innovation in the States: A Diffusion Study. *American Political Science Review*, *67*(4), 1174–1185. https://doi.org/10.2307/1956539

Graziadei, M. (2009). Legal Transplants and the Frontiers of Legal Knowledge. *Theoretical Inquiries in Law*, *10*(2), 723–744.

Graziadei, M. (2019). Comparative Law, Transplants, and Receptions. In M. Reimann & R. Zimmermann (Eds.), *The Oxford Handbook of Comparative Law* (pp. 441–473). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198810230.013.14

Greenleaf, G. (2009). Five years of the APEC Privacy Framework: Failure or promise? *Computer Law & Security Review*, *25*(1), 28–43. https://doi.org/10.1016/j.clsr.2008.12.002

Greenleaf, G. (2011). Global data privacy laws: 40 years of acceleration. *Privacy Laws & Business International Report*, *1*(112), 11–17.

Greenleaf, G. (2012). Independence of data privacy authorities (Part I): International standards. *Computer Law & Security Review*, *28*(1), 3–13. https://doi.org/10.1016/j.clsr.2011.12.001

Greenleaf, G. (2013). Sheherezade and the 101 Data Privacy Laws: Origins, Significance and Global Trajectories. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2280877

Greenleaf, G. (2014). Asian Data Privacy Laws—Trajectories, Lessons, and Optimism. In *Asian Data Privacy Laws: Trade & Human Rights Perspectives* (pp. 552–564). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199679669.001.0001

Greenleaf, G. (2015a). Global data privacy laws 2015: 109 countries, with European laws now a minority. *Privacy Laws & Business International Report*, *133*(1), 1–7.

Greenleaf, G. (2015b). Global data privacy laws 2015: Data privacy authorities and their organizations. *Privacy Laws & Business International Report*, *134*(1), 16–19.

Greenleaf, G. (2021). Global Tables of Data Privacy Laws and Bills. *Privacy Laws & Business International Report*, *169*(1), 6–19.

Greenleaf, G. (2022). Now 157 countries: Twelve data privacy laws in 2021/2022. *Privacy Laws & Business International Report*, *176*(1), 3–8.

Grimmer, J., & Stewart, B. M. (2013). Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. *Political Analysis*, *21*(3), 267–297. https://doi.org/10.1093/pan/mps028

Haas, P. M. (1992). Introduction: Epistemic communities and international policy coordination. *International Organization*, *46*(1), 1–35. https://doi.org/10.1017/S0020818300001442

Hafner-Burton, E. M., & Montgomery, A. H. (2011). Globalization and the Social Power Politics of International Economic Networks. In M. Kahler (Ed.), *Networked Politics: Agency, Power, and Governance* (pp. 23–42). Cornell University Press.

Harris, P. (1982). Structural change in the communication of precedent among state supreme courts, 1870–1970. *Social Networks*, *4*(3), 201–212. https://doi.org/10.1016/0378-8733(82)90021-1

Hartzog, W. (2017). The Inadequate, Invaluable Fair Information Practices. *Maryland Law Review*, *76*(4), 952–983.

Heclo, H. (1974). *Modern Social Politics in Britain and Sweden*. Yale University Press.

Heinz, J. P., & Laumann, E. O. (1982). *Chicago Lawyers: The Social Structure of the Bar*. Russel Sage Foundation.

Heslop-Harrison, J. S., & Schwarzacher, T. (2007). Domestication, Genomics and the Future for Banana. *Annals of Botany*, *100*(5), 1073–1084. https://doi.org/10.1093/aob/mcm191

Hirsch, D. D. (2013). Going Dutch? Collaborative Dutch Privacy Regulation and the Lessons it Holds for U.S. Privacy Law. *Michigan State Law Review*, *2013*(1), 83–166.

Hodgson, D. L. (2017). *Gender, Justice, and the Problem of Culture: From Customary Law to Human Rights in Tanzania*. Indiana University Press.

Hondius, F. (1975). *Emerging Data Protection in Europe*. North-Holland Publishing Company.

Horner, K. (2009). Regimenting language, mobility and citizenship in Luxembourg. In *Language Testing, Migration and Citizenship: Cross-National Perspectives on Integration Regimes* (pp. 148–169). A&C Black.

Humphries, M. D., & Gurney, K. (2008). Network 'Small-World-Ness': A Quantitative Method for Determining Canonical Network Equivalence. *PLoS ONE*, *3*(4), e0002051. https://doi.org/10.1371/journal.pone.0002051

Jean-Baptiste, R. (2008). "These Laws Should be Made by Us": Customary Marriage Law, Codification and Political Authority in Twentieth-Century Colonial Gabon. *Journal of African History*, *49*(2), 217–240.

Johns, F., & Joyce, D. (2014). Beyond Privacy. Is prevailing legal debate too analog for a digital age? *Human Rights Defender*, *23*(3), 24–26.

Josua, M. (2021). What Drives Diffusion? Anti-Terrorism Legislation in the Arab Middle East and North Africa. *Journal of Global Security Studies*, *6*(3), ogaa049. https://doi.org/10.1093/jogss/ogaa049

Jung, C. Y., & Joo, H. K. (2023). Post-'Lee-Luda' personal information protection in Korea: Developer responsibility and autonomous AI governance. *International Data Privacy Law*, ipad006. https://doi.org/10.1093/idpl/ipad006

Kahn-Freund, O. (1974). On Uses and Misuses of Comparative Law. *Modern Law Review*, *37*(1), 1–27.

Kalantry, S. (2020). Reverse Legal Transplants. *North Carolina Law Review*, *99*(1), 1–48.

Keohane, R. O., & Nye, J. S. (2001). *Power and Interdependence* (Third). Longman.

Kimbel-Sannit, A. (2023, May 12). How the Montana GOP supermajority wielded its power. *Montana Free Press*.

Kinikoglu, B. (2023). Implementing a new data protection law: Lessons from the Turkish experience. *International Data Privacy Law*, *13*(1), 25–46. https://doi.org/10.1093/idpl/ipad001

Kirby, M. D. (1980). Transborder Data Flows and the “Basic Rules” of Data Privacy. *Stanford Journal of International Law*, *16*(1), 27–66.

Kirby, M. D. (1991). Legal Aspects of Transborder Data Flows. *Computer/Law Journal*, *11*(1), 233–246.

Kokott, J., & Sobotta, C. (2013). The distinction between privacy and data protection in the jurisprudence of the CJEU and the ECtHR. *International Data Privacy Law*, *3*(4), 222–228.

Koniaris, M., Anagnostopoulos, I., & Vassiliou, Y. (2014). Legislation as a complex network: Modelling and analysis. *Frontiers in Artificial Intelligence and Applications*, *271*(1), 143–152.

Koops, B.-J. (2014). The trouble with European data protection law. *International Data Privacy Law*, *4*(4), 250–261.

Kuner, C. (2015). Extraterritoriality and regulation of international data transfers in EU data protection law. *International Data Privacy Law*, *5*(4), 235–245. https://doi.org/10.1093/idpl/ipv019

Kuner, C., Bygrave, L. A., Docksey, C., & Drechsler, L. (2020). *The EU General Data Protection Regulation (GDPR): A commentary*. Oxford University Press.

Kuner, C., Cate, F. H., Millard, C., Svantesson, D. J. B., & Lynskey, O. (2015). Internet Balkanization gathers pace: Is privacy the real driver? *International Data Privacy Law*, *5*(1), 1–2. https://doi.org/10.1093/idpl/ipu032

Lando, O. (2006). On Legislative Style and Structure. *Juridica International*, *11*, 13–20.

Laurer, M., & Seidl, T. (2021). Regulating the European Data-Driven Economy: A Case Study on the General Data Protection Regulation. *Policy & Internet*, *13*(2), 257–277. https://doi.org/10.1002/poi3.246

Law, D. S., & Versteeg, M. (2013). Sham Constitutions. *University of California Law Review*, *101*(4), 863–952.

Lazer, D. (2005). Regulatory Capitalism as a Networked Order: The International System as an Informational Network. *The ANNALS of the American Academy of Political and Social Science*, *598*(1), 52–66. https://doi.org/10.1177/0002716204272590

Lazer, D. (2006). Global and Domestic Governance: Modes of Interdependence in Regulatory Policymaking. *European Law Journal*, *12*(4), 455–468.

Lee-Makiyama, H. (2014). The Political Economy of Data: EU Privacy Regulation and the International Redistribution of Its Costs. In L. Floridi (Ed.), *Protection of Information and the Right to Privacy—A New Equilibrium* (Vol. 17). Springer.

Legrand, P. (1997). The Impossibility of Legal Transplants. *Maastricht Journal of European & Comparative Law*, *4*(2), 111–124.

Leichter, H. (1986). Saving Lives and Protecting Liberty: A Comparative Study of the Seat-Belt Debate. *Journal of Health Politics, Policy and Law*, *11*(2), 323–344. https://doi.org/10.1215/03616878-11-2-323

Linder, F., Desmarais, B., Burgess, M., & Giraudy, E. (2020). Text as Policy: Measuring Policy Similarity through Bill Text Reuse. *Policy Studies Journal*, *48*(2), 546–574. https://doi.org/10.1111/psj.12257

Lynskey, O. (2015). *The Foundations of EU Data Protection Law*. Oxford University Press.

Maximilian Schrems v. Data Protection Commissioner [Schrems I], Case C-362/14 (Court of Justice of the European Union 2015).

May, P. J. (1992). Policy Learning and Failure. *Journal of Public Policy*, *12*(4), 331–354. https://doi.org/10.1017/S0143814X00005602

Mayer-Schönberger, V. (1997). Generational Development of Data Protection in Europe. In M. Rotenberg & P. E. Agre (Eds.), *Technology and Privacy: The New Landscape* (pp. 219–241). MIT Press.

Mayer-Schönberger, V. (2021). Paradigm Shift. *Computer Law & Security Review*, *40*(1), 1–3.

Mayer-Schönberger, V., & Gasser, U. (2024, September 16). A Realist Perspective on AI Regulation. *Foreign Policy*. https://foreignpolicy.com/2024/09/16/realist-perspective-ai-regulation/

Mayer-Schönberger, V., & Lazer, D. (2002). Governing Networks. *Brooklyn Journal of International Law*, *27*(3), 819–851.

Mendez, F., & Mendez, M. (2010). Comparing Privacy Regimes: Federal Theory and the Politics of Privacy Regulation in the European Union and the United States. *Publius*, *40*(4), 617–645.

Miller, A. R. (1971). *Assault on Privacy: Computers, Data Banks and Dossiers*. University of Michigan Press.

Miller, J. A. (2003). A Typology of Legal Transplants. *American Journal of Comparative Law*, *51*(4), 839–886.

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., & Alon, U. (2002). Network Motifs: Simple Building Blocks of Complex Networks. *Science*, *298*(5594), 824–827. https://doi.org/10.1126/science.298.5594.824

Moerel, L. (2011). The long arm of EU data protection law: Does the Data Protection Directive apply to processing of personal data of EU citizens by websites worldwide? *International Data Privacy Law*, *1*(1), 28–46. https://doi.org/10.1093/idpl/ipq004

Moodley, K., Hernandez Serrano, P. V., Van Dijck, G., & Dumontier, M. (2019). Similarity and Relevance of Court Decisions: A Computational Study on CJEU Cases. In M. Araszkiewicz & V. Rodriguez Doncel (Eds.), *Legal Knowledge and Information Systems: JURIX 2019* (pp. 63–72). IOS Press.

Moyson, S., Scholten, P., & Weible, C. M. (2017). Policy learning and policy change: Theorizing their relations from different perspectives. *Policy and Society*, *36*(2), 161–177. https://doi.org/10.1080/14494035.2017.1331879

Neumayer, E., & Plumper, T. (2012). Conditional Spatial Policy Dependence: Theory and Model Specification. *Comparative Political Studies*, *45*(7), 819–849.

Newman, A. L. (2008). *Protectors of Privacy: Regulating Personal Data in the Global Economy*. Cornell University Press.

Nissenbaum, H. (2010). *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Stanford University Press.

Nissenbaum, H. (2017). Protecting Privacy in an Information Age: The Problem of Privacy in Public. In K. Miller & M. Taddeo (Eds.), *The Ethics of Information Technologies*. Routledge.

OECD. (1980). *Recommendation Concerning Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data*.

Ohm, P. (2010). Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization. *UCLA Law Review*, *57*(1), 1701–1777.

Ojwang, J. (2024, October 17). Tanzania under pressure to address rights violations ahead of local elections. *Jurist*.

Olson, M. (1965). *The Logic of Collective Action*. Yale University Press.

Orji, U. J. (2017). Regionalizing data protection law: A discourse on the status and implementation of the ECOWAS Data Protection Act. *International Data Privacy Law*, *7*(3), 179–189. https://doi.org/10.1093/idpl/ipx013

Paal, B. P. (2022). Artificial Intelligence as a Challenge for Data Protection Law: And Vice Versa. In S. Voeneky, P. Kellmeyer, O. Mueller, & W. Burgard (Eds.), *The Cambridge Handbook of Responsible Artificial Intelligence* (1st ed., pp. 290–308). Cambridge University Press. https://doi.org/10.1017/9781009207898.023

Palladino, N. (2021). The role of epistemic communities in the "constitutionalization" of internet governance: The example of the European Commission High-Level Expert Group on Artificial Intelligence. *Telecommunications Policy*, *45*(6), 102149. https://doi.org/10.1016/j.telpol.2021.102149

Panzavolta, M. (2016). Of hearsay and beyond: Is the Italian criminal justice system an adversarial system? *The International Journal of Human Rights*, *20*(5), 617–633. https://doi.org/10.1080/13642987.2016.1162409

Petkova, B. (2016). The Safeguards of Privacy Federalism. *Lewis and Clark Law Review*, *20*(2), 595–645.

Pisa, M., & Nwankwo, U. (2021, August). *Are Current Models of Data Protection Fit for Purpose? Understanding the Consequences for Economic Development: Roundtable Summary*. Center for Global Development. https://www.cgdev.org/sites/default/files/are-current-models-data-protection-fit-purpose-understanding-consequences-economic.pdf

Puchniak, D. W., & Luh Luh, L. (2017). Independent Directors in Singapore: Puzzling Compliance Requiring Explanation. *American Journal of Comparative Law*, *65*(2), 265–334.

Puig, S. (2018). Network analysis and the sociology of international law. In M. Hirsch & A. Lang (Eds.), *Research Handbook on the Sociology of International Law*. Edward Elgar Publishing. https://doi.org/10.4337/9781783474493.00022

Putnam, R. D. (1988). Diplomacy and domestic politics: The logic of two-level games. *International Organization*, *41*(3), 427–460.

Rajan, A. (2021, October 13). *Information commissioner Elizabeth Denham: How to be a pro-active regulator*. BBC. https://www.bbc.co.uk/news/entertainment-arts-58903479

Regan, P. M. (1995). *Legislating Privacy: Technology, Social Values, and Public Policy*. University of North Carolina Press.

Reidenberg, J. R. (2000). Resolving Conflicting International Data Privacy Rules in Cyberspace. *Stanford Law Review*, *52*(5), 1315. https://doi.org/10.2307/1229516

Resta, G. (2014). Systematic government access to private-sector data in Italy. *International Data Privacy Law*, *4*(1), 12–20. https://doi.org/10.1093/idpl/ipt029

Reviglio, U., & Alunge, R. (2020). “I Am Datafied Because We Are Datafied”: An Ubuntu Perspective on (Relational) Privacy. *Philosophy & Technology*, *33*(4), 595–612. https://doi.org/10.1007/s13347-020-00407-6

Riccardi, J. L. (1983). The German Federal Data Protection Act of 1977: Protecting the Right to Privacy? *Boston College International and Comparative Law Review*, *61*(1), 243–271.

Richie, G. E. (2010). *The Role of the Epistemic Community in Influencing Privacy Legislation: The United States and the European Union*. University of Denver.

Roe, M. J. (2005). Delaware’s Politics. *Harvard Law Review*, *118*(8), 2491–2543.

Romano, R. (1985). Law as a Product: Some Pieces of the Incorporation Puzzle. *The Journal of Law, Economics, and Organization*, *1*(2), 225–283.

Rose, R. (1991). What is lesson-drawing? *Journal of Public Policy*, *11*(1), 3–30.

Rotenberg, J. (2020). Privacy before Trade: Assessing the WTO-Consistency of Privacy-Based Cross-Border Data Flow Restrictions. *University of Miami International and Comparative Law Review*, *28*(1), 91–1200.

Rustad, M. L., & Koenig, T. H. (2019). Towards a Global Data Privacy Standard. *Florida Law Review*, *71*(2), 365–454.

Rwezaura, B. (1994). Tanzania: Building a New Family Law out of a Plural Legal System. *University of Louisville Journal of Family Law*, *33*(2), 523–540.

Sabatier, P. A. (1987). Knowledge, Policy-Oriented Learning, and Policy Change: An Advocacy Coalition Framework. *Knowledge*, *8*(4), 649–692. https://doi.org/10.1177/016402598700800400 5

Sacco, R. (1991). Legal Formants: A Dynamic Approach to Comparative Law (Installment I of II). *The American Journal of Comparative Law*, *39*(1), 1. https://doi.org/10.2307/840669

Sanga, S. (2020). Network Effects in Corporate Governance. *The Journal of Law and Economics*, *63*(1), 1–41. https://doi.org/10.1086/706191

Schmidt, R. (2018). Integration, Networks and the Global Order. In *Regulatory Integration Across Borders: Public-Private Cooperation in Transnational Regulation* (pp. 71–104). Cambridge University Press.

Schwartz, P. M. (2013). The EU-U.S. Privacy Collision: A Turn to Institutions and Procedures. *Harvard Law Review*, *126*(7).

Schwartz, P. M. (2015). The Value of Privacy Federalism. In B. Roessler & D. Mokrosinska (Eds.), *Social Dimensions of Privacy; Interdisciplinary Perspectives*. Cambridge University Press.

Schwartz, P. M. (2019). *Global Data Privacy: The EU Way*. *94*(4), 771–818.

Schwartz, P. M., & Peifer, K.-N. (2017). Transatlantic Data Privacy Law. *The Georgetown Law Journal*, *106*(1), 115–180.

Schwartz, P. M., & Solove, D. J. (2011). *The PII Problem: Privacy and a New Concept of Personally Identifiable Information*. *86*(6), 1814–1894.

Shaffer, R., & Elkins, Z. (2019). *On Measuring Textual Similarity*. Rbshaffer.Github.Io. https://rbshaffer.github.io/_includes/On_Measuring_Textual_Similarity.pdf

Sharpe, M. (2017, May 23). *UK Political Manifestos*. DogDogFish: Data Science, amongst Other Things. https://dogdogfish.com/python-2/uk-political-manifestos/

Shipan, C. R., & Volden, C. (2008). The Mechanisms of Policy Diffusion. *American Journal of Political Science*, *52*(4), 840–857. https://doi.org/10.1111/j.1540-5907.2008.00346.x

Shur-Ofry, M., Fibich, G., & Green, S. (2019). The Diffusion of Legal Innovation—Insights from Mathematical Modeling. *Cornell International Law Journal*, *52*(2), 313–350.

Siems, M. (2018). Malicious legal transplants. *Legal Studies Research Paper Series*, *38*(1), 103–119.

Siems, M. (2022). *Comparative Law* (Third). Cambridge University Press.

Siems, M. (2023). A Network Analysis of Judicial Cross-Citations in Europe. *Law & Social Inquiry*, *48*(3), 881–905. https://doi.org/10.1017/lsi.2022.22

Simitis, S. (1987). Reviewing Privacy in an Information Society. *University of Pennsylvania Law Review*, *135*(3), 707–746.

Simitis, S., & Dammann, U. (1977). *Data Protection Legislation: An International Documentation*. A. Metzner.

Singh Grewal, D. (2008). *Network Power*. Yale University Press.

Slaughter, A.-M. (2017). The Accountability of Government Networks. In P. S. Berman (Ed.), *The Globalization of International Law* (First edition). Routledge.

Smith, J. (2007). Federalism and Democratic Accountability. In J. DeBardeleben & A. Hurrelmann (Eds.), *Democratic Dilemmas of Multilevel Governance* (pp. 38–58). Palgrave Macmillan UK. https://doi.org/10.1057/9780230591783_3

Solove, D. J. (2002). Conceptualizing Privacy. *University of California Law Review*, *90*(4), 1087–1155.

Solove, D. J. (2006). A Taxonomy of Privacy. *University of Pennsylvania Law Review*, *154*(3), 477–564.

Spamann, H. (2009). Contemporary Legal Transplants: Legal Families and the Diffusion of (Corporate) Law. *Brigham Young University Law Review*, *2009*(6), 1813–1878.

Spirling, A. (2012). U.S. Treaty Making with American Indians: Institutional Change and Relative Power, 1784-1911. *American Journal of Political Science*, *56*(1), 84–97.

Stadlen, G. (1979). Survey of National Data Protection Legislation. *Computer Networks*, *3*(3), 174–186.

Stein, A. A. (2009). Neoliberal Institutionalism. In C. Reus-Smit & D. Snidal (Eds.), *The Oxford Handbook of International Relations* (1st ed., pp. 201–221). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199219322.003.0011

Stephan, P. B. (2023). *The World Crisis and International Law: The Knowledge Economy and the Battle for the Future*. Cambridge University Press.

Strandburg, K. J., Viljoen, S., & Nissenbaum, H. F. (2024). The Great Regulatory Dodge. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4852257

Sudarwanto, A. S., & Kharisma, D. B. B. (2022). Comparative study of personal data protection regulations in Indonesia, Hong Kong and Malaysia. *Journal of Financial Crime*, *29*(4), 1443–1457. https://doi.org/10.1108/JFC-09-2021-0193

Sulis, E., Bryn Humphreys, L., Audrito, D., & Di Caro, L. (2021). Exploiting Textual Similarity Techniques in Harmonization of Laws. In S. Bandini, F. Gasparini, V. Mascardi, M. Palmonari, & G. Vizzari (Eds.), *AIxIA 2021—Advances in Artificial Intelligence* (pp. 185–197). Springer.

Sullivan, C. (2019). EU GDPR or APEC CBPR? A comparative analysis of the approach of the EU and APEC to cross border data transfers and protection of personal data in the IoT era. *Computer Law & Security Review*, *35*(4), 380–397. https://doi.org/10.1016/j.clsr.2019.05.004

Susser, D. (2019). Notice After Notice-and-Consent: Why Privacy Disclosures Are Valuable Even If Consent Frameworks Aren't. *Journal of Information Policy*, *9*, 148–173. https://doi.org/10.5325/jinfopoli.9.2019.0148

Svantesson, D. J. B. (2013). *Extraterritoriality in Data Privacy Law*. Narayana Press.

Swire, P. (2013). The Second Wave of Global Privacy Protection: Symposium Introduction. *Ohio State Law Journal*, *74*(6), 841–852.

Tam Cho, W. K., & Fowler, J. H. (2010). Legislative Success in a Small World: Social Network Analysis and the Dynamics of Congressional Legislation. *The Journal of Politics*, *72*(1), 124–135. https://doi.org/10.1017/S002238160999051X

Tavani, H. T. (2007). Philosophical Theories of Privacy: Implications for an Adequate Online Privacy Policy. *Metaphilosophy*, *38*(1), 1–22.

Telesford, Q. K., Joyce, K. E., Hayasaka, S., Burdette, J. H., & Laurienti, P. J. (2011). The Ubiquity of Small-World Networks. *Brain Connectivity*, *1*(5), 367–375. https://doi.org/10.1089/brain.2011.0038

Teubner, G. (1998). Legal Irritants: Good Faith in British Law or How Unifying Law Ends up in New Divergences. *Modern Law Review*, *61*(1), 11–32.

The Privacy Advisor. (2019, May 2). *50k Members: A landmark for the IAPP and global privacy*. Iapp. https://iapp.org/news/a/50k-members-a-landmark-for-iapp-and-global-privacy/

Thouvenin, F. (2021). Informational Self-Determination: A Convincing Rationale for Data Protection Law? *Journal of Intellectual Property, Information Technology and Electronic Commerce Law*, *12*(4), 246–256.

Twining, W. (2004). Diffusion of Law: A Global Perspective. *The Journal of Legal Pluralism and Unofficial Law*, *36*(49), 1–45. https://doi.org/10.1080/07329113.2004.10756300

Twining, W. (2005). Social Science and Diffusion of Law. *Journal of Law and Society*, *32*(2), 203–240.

Vatanparast, R. (2020). Designed to Serve Mandkind? The Politics of the GDPR as a Global Standard and the Limits of Privacy. *Heidelberg Journal of International Law*, *80*(4), 819–845.

Viljoen, S. (2021). *A Relational Theory of Data Governance*. *131*(2), 573–654.

Vogel, D. (1997). *Trading Up: Consumer and Environmental Regulation in a Global Economy*. Harvard University Press.

Vogel, D. (2018). *California Greenin': How the Golden State Became an Environmental Leader*. Princeton University Press.

Wachovia Bank, N.A. v. Schmidt, 546 United States Reports 303 (United States Supreme Court 2006).

Waldman, A. E. (2019). Privacy Law's False Promise. *Washington University Law Review*, *97*(3), 773–834.

Walter, A. (2023, July 11). Privacy Shield 2.0: EU-US Data Privacy Framework adopted. *Out-Law | Your Daily Need to Know | Pinsent Masons*. https://www.pinsentmasons.com/out-law/news/privacy-shield-2-eu-us-adopted

Wang, J., & Dong, Y. (2020). Measurement of Text Similarity: A Survey. *Information*, *11*(9), 421. https://doi.org/10.3390/info11090421

Watson, A. (1976). Legal Transplants and Law Reform. *Law Quarterly Review*, *92*, 79–84.

Watson, A. (1993). *Legal Transplants: An Approach to Comparative Law* (Second, Vol. 92). University of Georgia Press.

Watson, A. (1995). From legal transplants to legal formants. *American Journal of Comparative Law*, *43*(3), 469–476.

Watts, D. J. (2004). The "New" Science of Networks. *Annual Review of Sociology*, *30*(1), 243–270. https://doi.org/10.1146/annurev.soc.30.020404.104342

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of "small-world" networks. *Nature*, *393*(4), 440–442.

Weitzenboeck, E. M., Lison, P., Cyndecka, M., & Langford, M. (2022). The GDPR and unstructured data: Is anonymization possible? *International Data Privacy Law*, *12*(3), 184–206. https://doi.org/10.1093/idpl/ipac008

Westin, A. F. (1967). *Privacy and Freedom*. Atheneum.

Whalen, R. (2016). Legal Networks: The Promises and Challenges of Legal Network Analysis. *Michigan State Law Review*, *2016*(2), 539–565.

Whitman, J. Q. (2004). The Two Western Cultures of Privacy: Dignity versus Liberty. *Yale Law Journal*, *113*(6), 1151–1222.

Wilkerson, J., & Casas, A. (2017). Large-Scale Computerized Text Analysis in Political Science: Opportunities and Challenges. *Annual Review of Political Science*, *20*(1), 529–544. https://doi.org/10.1146/annurev-polisci-052615-025542

Wilkerson, J., Smith, D., & Stramp, N. (2015). Tracing the flow of policy ideas in legislatures: A text reuse approach. *American Journal of Political Science*, *59*(4), 943–956.

Wilkinson, L., & Friendly, M. (2009). The History of the Cluster Heat Map. *The American Statistician*, *63*(2), 179–184. https://doi.org/10.1198/tas.2009.0033

Wohlforth, W. C. (2009). Realism. In C. Reus-Smit & D. Snidal (Eds.), *The Oxford Handbook of International Relations* (pp. 131–149). Oxford University Press.

Yang, L., & Yan, M. (2021). The Conceptual Barrier to Comparative Study and International Harmonisation of Data Protection Law. *Hong Kong Law Journal*, *51*(3), 917–950.

Zhou, K., Ethayarajh, K., Card, D., & Jurafsky, D. (2022). *Problems with Cosine as a Measure of Embedding Similarity for High Frequency Words*. https://doi.org/10.48550/ARXIV.2205.05092

Zuboff, S. (2015). Big other: Surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology*, *30*(1), 75–89.