Original Research Article

BIG DATA & SOCIETY

Big Data & Society
January-June: 1–17

DOI: 10.1177/2053951720982012
journals.sagepub.com/home/bds
SAGE

Data sovereignty: A review

Patrik Hummel, Matthias Braun, Max Tretter and Peter Dabrock

Abstract

New data-driven technologies yield benefits and potentials, but also confront different agents and stakeholders with challenges in retaining control over their data. Our goal in this study is to arrive at a clear picture of what is meant by data sovereignty in such problem settings. To this end, we review 341 publications and analyze the frequency of different notions such as *data* sovereignty, *digital* sovereignty, and *cyber* sovereignty. We go on to map *agents* they concern, in which *context* they appear, and which *values* they allude to. While our sample reveals a considerable degree of divergence and an occasional lack of clarity about intended meanings of data sovereignty, we propose a conceptual grid to systematize different dimensions and connotations. Each of them relates in some way to meaningful control, ownership, and other claims to data articulated by a variety of agents ranging from individuals to countries. Data sovereignty alludes to a nuanced mixture of normative concepts such as inclusive deliberation and recognition of the fundamental rights of data subjects.

Keywords

Digitization, privacy, deliberation, sovereignty, power, ownership

Introduction

In the age of digitization, dealing responsibly with data poses a dilemma: on the one hand, there is individually tangible and easily comprehensible added value of personal data processing by public and private-sector institutions. Examples include the personalization of products and services, the refinement of healthcare, availability of favorable insurance rates for individuals with low-risk profiles, and benefits made possible by the donation of sensitive personal data, e.g. in order to advance research and public health surveillance. On the other hand, there is the more or less abstract idea that individuals, specific groups, or communities should retain control over the handling of their data.

In order to address issues like these, recent years have seen an emergence of the notion of data sovereignty in debates on the development, implementation, and adjustment of new data-driven technologies and their infrastructures. Data is a timely subject matter, given that they mediate and steer extensive parts of our lifeworld. And sovereignty, understood, e.g. as the ability to issue authoritative claims, latching onto domestic institutional arrangements, international regimes, and the practices of other states (Krasner, 1988), appears as a fruitful category to apply to data. At the same time, it often remains implicit and contentious what exactly data sovereignty means, and also what, if anything, distinguishes it from other notions of sovereignty, e.g. cyber sovereignty, internet sovereignty, or "[t]he fight for digital sovereignty" and its entanglement with "analogue", "national", "socio-political" (Floridi, 2020) sovereignty.

Recently, excellent review articles on sovereignty and the digital have appeared (Baezner and Robin, 2018; Couture and Toupin, 2019). These reviews underline the rich and diverging ways in which sovereignty has been treated in current governance discourses, and provide illuminating narrative reviews of selected materials. According to these reviews, data sovereignty involves, or can be identified with, the control of data flows via national jurisdiction. However, as these studies themselves indicate, further systematic analyses are needed. Indeed, questions about generalizability loom once we consider tensions between definitions like these and other calls for data sovereignty that concern a

Institute for Systematic Theology, Friedrich-Alexander-Universitat Erlangen-Nurnberg, Erlangen, Germany

Corresponding author:
Patrik Hummel, Institute for Systematic Theology, Friedrich-Alexander-Universitat Erlangen-Nurnberg, Kochstraße 6, Erlangen 91054, Germany.
Email: patrik.hummel@fau.de

broader variety of themes, including the authority of national governments over data stored in domestic or foreign clouds (Irion, 2012), but also research and surveillance vis-à-vis Indigenous data sovereignty (Kukutai and Taylor, 2016), and patient data sovereignty over health data (German Ethics Council, 2017).

Thus, there is a need to investigate systematically what data sovereignty is intended to encompass. Authors might be talking past each other or make vague policy demands if they call for or dispute data sovereignty without being explicit about which of the various potential connotations are intended. Such clarification would be helpful to researchers, policy makers, activists, and others who consider availing themselves of the notion and leveraging it towards their ends.

The present study seeks to fill this gap by providing an in-depth review of data sovereignty as it is used in academic journal publications. It is motivated by the observation that as illustrated by the foregoing examples, i.e. national data sovereignty in cloud computing, Indigenous data sovereignty, and patient data sovereignty, there is variance with regard to how data sovereignty is understood. In view of different connotations, claims, and objectives, the question arises what unites these different uses. Neither confronting pre-theoretical intuitions nor taking them for granted, our study seeks to provide data that could serve to evaluate these intuitions, to illuminate specifics, and to potentially uncover aspects of data sovereignty that might have escaped awareness. Academic literature *per se* is not in a privileged position to expound the notion, but presumably reflects understandings from different domains, and will be a reference point once disagreements on data sovereignty surface, e.g. in the political sphere.

While pursuing this project, we bracket our own account of the subject matter as published in previous papers (e.g. Hummel et al., 2018). Neither do we intend to construct, re-engineer, or ameliorate the notion. Instead, our objective is to map different understandings of data sovereignty, the agents it concerns, and the normative concepts it embeds, and to sharpen its contours by contrasting it with other notions of sovereignty.

Methods

Guided by frameworks on conducting and reporting on systematic reviews (Moher et al., 2009; Strech and Sofaer, 2012; Tranfield et al., 2003), we began by formulating a search strategy (Table 1 in the Supplemental Appendix) for addressing the review question in which meanings of data sovereignty are used or presupposed in journal publications. Besides the term "data sovereignty", the strategy includes the three cognate notions that appeared relevant to us in light of cursory scoping searches: "digital sovereignty", "cyber sovereignty", and "virtual sovereignty". The review process began in August 2018, and the cutoff was November 2019.

Table 1. Notions and their number of occurrences in our sample.

	Data	Other	Cyber	Digital	Internet	Total
#	680	314	213	175	110	1492
%	45.6	21.0	14.3	11.7	7.4	

The search strategy returned 705 publications. After removing duplicates (76) and excluding items due to missing full-text (27), 602 full-text publications were screened (Figure 1). A paper was included in our review if it mentioned a notion of sovereignty in connection with the digital, e.g. data, digitization, information and communications technology, or digital infrastructures. Due to the composition of our research team, only publications in English or German were included. We did not assess the quality of discussions or arguments, given our goal to map, rather than to evaluate, uses of the notions in the search strategy. A paper was excluded if it did not discuss sovereignty, or if sovereignty (or a specific kind thereof, e.g. state sovereignty, food sovereignty) was not raised in connection with the digital.

Included papers were read and analyzed with a focus on the passages mentioning sovereignty. Our goal was to develop categories for comparisons of passages, ideally arriving at a scheme that is systematic while allowing for "exploration, discovery and development" (Tranfield et al., 2003: 215) and without presupposing a particular theory of data sovereignty. To this end, we proceeded inductively by observing the data and across several iterations capturing similarities and differences that caught our attention. In this process, a set of five *questions* emerged on which passages differed:

1. *Notion*: which *notion* of sovereignty was mentioned?
2. *Agents*: which entities are involved in ascriptions of sovereignty in the respective passage?
3. *Contexts*: what is the broader domain and/or topic that is the background of a mention of sovereignty?
4. *Values*: which concepts from normative theorizing (e.g. ethics, legal theory) are mentioned in connection with sovereignty?
5. *Content*: how was the *notion* used: descriptively, as highlighting a challenge, or as a target to attain?

While comparing the presuppositions of passages on these five questions, we developed an inductive conceptual scheme (Table 2 in the Supplemental Appendix), which we refined through several iterations of adding, merging, or splitting categories that capture how these five *questions* were treated in a passage. In the

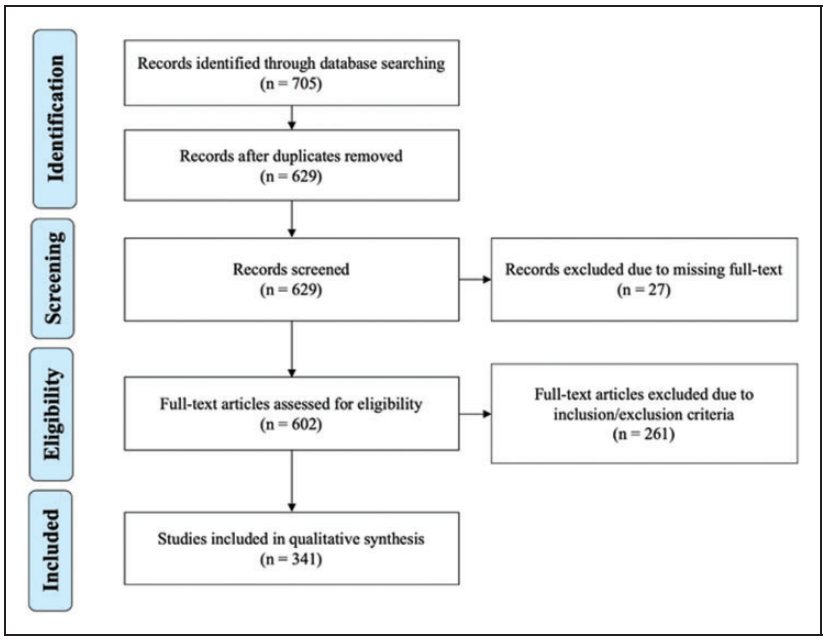


Figure 1. Screening process. Adapted from Moher et al. (2009).

refinement process, new passages were examined in light of intermediate categories that had emerged while addressing the *questions* for previous passages. Passages were categorized by means of the qualitative content analysis software Atlas.Ti. Amongst the *notions*, we distinguished *data*, *cyber*, *digital*, *internet* sovereignty, and *other*s. Regarding *questions* 2–5, the scheme captures a range of 4–16 categories per *question*. These categories and their definitions are provided in Table 2 in the Supplemental Appendix.

In the following, we use italics to refer to concepts figuring in the *questions* and our inductive scheme. We use underlining to highlight particular words indicating what we are illustrating by means of a quoted passage. As one example, consider the following passage:

> To protect the privacy of health data, the German eHC [electronic health card] system defines basic security requirements: The authentication, authorization, and audit mechanisms have to be chosen in a way that the data sovereignty of the insured party can be taken for granted. This means that by using the eHC the insured person can solely control the access to the health data. Technically, this is realized by a hybrid encryption mechanism: The health data are encrypted with a random symmetric key before they are uploaded to the EHR server, and this key is encrypted with the public key of the patient's eHC. (Winandy, 2011: 197)

In this passage, the *notion* is *data sovereignty*. Given its focus on the specifics of the German eHC system, the passage was coded with two *context* classifiers: *clinical practice* and *IT architecture*. The *agent* mentioned is the *patient*, and the *values* are *control and power* as well as *privacy*. With regards to the *content*, *data sovereignty* is presented as a target, and the passage provides a *management strategy* to attain it.

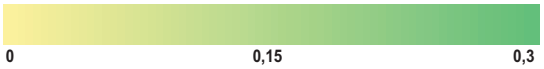
Some of the concepts in the categories (especially the *values*, but also *agents* like *citizen*) hardly admit of uncontested, theory-neutral necessary and sufficient conditions for their application. Moreover, depending on how one makes them precise, some categories can overlap (e.g. amongst the *values: autonomy*, *control and power*, and *privacy*). We addressed both challenges as follows: a passage was coded as pertaining to the category if the category was mentioned explicitly (e.g. "privacy", "control"), *or* if implicit or explicit references to components of the definitions outlined in Table 2 in the Supplemental Appendix were made. Moreover, as the example shows, passages could be coded with more than one category per *question*. For the vast majority of passages, at least one category per *question* could be assigned.

At the conclusion of the coding process, 1433 text passages across the 341 included publications were coded. We then used Atlas.Ti's analysis tools to compile the c-coefficients for category pairs from our inductive scheme. The c-coefficient is a measure for the co-occurrences (i.e. being mentioned in the same passage) between two categories (or "codes" in Atlas.Ti terminology):

$$c = \frac{\mathrm{n12}}{\mathrm{n1} + \mathrm{n2} - \mathrm{n12}}$$

Table 2. Heat map of the c-coefficients (co-occurrences) between notions and agents. The color coding reflects the c-coefficient.

notion-agent	data sov.(680)	other (314)	cyber sov. (213)	digital sov. (175)	internet sov. (110)
countries (549)	0,08 (96)	0,17 (128)	0,28 (169)	0,18 (109)	0,16 (91)
Indigenous population (217)	0,24 (176)	0,09 (42)	0 (0)	0,02 (6)	0 (0)
user/consumer (160)	0,14 (105)	0,1 (43)	0,01 (5)	0,03 (10)	0,01 (3)
private-sector organizations (146)	0,14 (104)	0,02 (11)	0,03 (11)	0,06 (17)	0,02 (4)
governmental organizations (118)	0,07 (54)	0,07 (27)	0,04 (13)	0,09 (23)	0,01 (3)
non-governmental organizations (88)	0,08(56)	0,06 (22)	0,02 (6)	0,04 (9)	0,01 (1)
expert/professional (76)	0,05 (36)	0,08 (28)	0,03 (8)	0,03 (7)	0,03 (5)
societies (73)	0,03 (22)	0,12 (40)	0,03 (7)	0,02 (6)	0 (0)
citizen (62)	0,04 (31)	0,06 (21)	0 (1)	0,03 (8)	0,01 (2)
intergovernmental organizations (42)	0,02 (14)	0,03 (11)	0,03 (7)	0,02 (5)	0,03 (5)
patient (26)	0,03 (23)	0,01 (5)	0 (0)	0 (0)	0 (0)



with n12 referring to the number of co-occurrences of two categories, and n1 and n2 referring to their respective individual number of occurrences (Friese, 2019: 151–152). Note that n12 is smaller than or equal to the maximum of n1 or n2. If n1 is much smaller than n2 or *vice versa*, the c-coefficient is systematically deflated and becomes uninformative. We used the c-coefficient as a pointer in our analysis, but also made sure to consider this metric alongside absolute frequencies (number of times a concept appears) in order to not miss out on interesting connections.

Results

In the following, we structure our presentation of the results on how *data sovereignty* is used in our sample along the five *questions* just outlined. In order to illuminate *data sovereignty*, we also consider the other

Table 3. Heat map of the c-coefficients (co-occurrences) between notions and contexts. The color coding reflects the c-coefficient.

notion-context	data sov.(680)	other (314)	cyber sov. (213)	digital sov. (175)	internet sov. (110)
legislation (380)	0,2 (173)	0,15 (88)	0,11 (60)	0,07 (38)	0,09 (42)
IT architecture (362)	0,26 (215)	0,08 (49)	0,06 (30)	0,13 (61)	0,05 (21)
research (208)	0,13 (105)	0,17 (77)	0,04 (16)	0,03 (11)	0,03 (9)
societal discourse and advocacy (154)	0,11 (86)	0,09 (38)	0,03 (10)	0,05 (17)	0,05 (12)
defense (138)	0,01 (8)	0,08 (35)	0,21 (60)	0,15 (42)	0,02 (6)
international relations (122)	0,02 (15)	0,09 (37)	0,14 (41)	0,04 (12)	0,12 (24)
business and economy (98)	0,06 (45)	0,05 (21)	0,03 (8)	0,06 (15)	0,05 (10)
surveillance (84)	0,03 (23)	0,06 (21)	0,11 (30)	0,06 (14)	0,04 (8)
education and capacity-building (73)	0,07 (47)	0,05 (17)	0,01 (2)	0,02 (4)	0,02 (3)
public administration (50)	0,06 (39)	0,01 (5)	0,01 (3)	0 (1)	0,02 (3)
clinical practice (27)	0,02 (15)	0,02 (7)	0,01 (3)	0,02 (3)	0 (0)
soft law (26)	0,04 (27)	0,01 (2)	0 (0)	0 (0)	0 (0)

notions of sovereignty and how they contrast with *data sovereignty*. As will become clear, a range of different approaches and understandings of *data sovereignty* can be distinguished.

Notions

Table 1 presents the frequency of the different *notions* in our sample. Tables 2 to 4 provide co-occurrences

between the *notions* and the categories in our inductive scheme. *Data sovereignty* has the highest frequency (680 or 45.6% of 1492 mentions). This is a noteworthy result about the salience of the notion, given that *data sovereignty* was not privileged in the search strategy relative to *digital*, *cyber*, and *virtual sovereignty* (Table 1 in the Supplemental Appendix). The category *Other* comprises further notions of sovereignty. Specifically, the five most frequent items in this category are sovereignty in general, i.e. without further specification (53), national or country sovereignty (49), genomic sovereignty (46), Indigenous or tribal sovereignty (35), and technological sovereignty (27). *Virtual* sovereignty was part of the search strategy, but appeared in only eight passages. Given this low number, we did not treat it separately, but included it amongst *Other*. In contrast, *internet sovereignty* was not explicitly part of the search strategy, but still appeared in 110 passages, which is why we captured it as an extra *notion* in our analysis.

Tables 2 to 4 provide an overview of how these *notions* co-occur with the categories of our conceptual scheme, i.e. the *contents*, *agents*, *contexts*, and *values* (all of these concepts as per the definitions provided in Table 3). Co-occurrence gives some indication on which concepts tend to be relevant when the respective *notion* is invoked. However, co-occurrence by itself does not indicate how exactly the *notion* relates to these co-occurring concepts. For example, co-occurrence between *data sovereignty* and the *agent* "governmental organization" is neutral on whether the latter is (supposed to be) *data sovereign* or whether it *determines* the *data sovereignty* of another agent. In order to better understand the specific connections in the co-occurrences, we take a closer look at their concrete modes of relatedness. In this process, we focus on some particularly salient co-occurrences, without claiming that categories co-occurring less frequently with *data sovereignty* have no interesting conceptual relation to this *notion*.

Content

As outlined in the "Methods" section and Table 2 in the Supplemental Appendix, we distinguished three kinds of *content* in the reviewed passages: *descriptions* of what sovereignty consists in, *challenges* to sovereignty, and *management strategies* for safeguarding and maintaining sovereignty. While the data does contain paradigmatic examples for these *contents*, it turned out that in other places these *contents* could not always be disentangled neatly, and a passage could be understood, e.g. as outlining a *challenge*, but also as implicitly providing a *description*. For this reason, we refrained from counting (and comparing the counts of) instances of these three *contents*. In order to address the question of our review, we now provide an overview on paradigmatic examples of these *contents* for *data sovereignty*.

To begin with, we turn to *descriptions* of what *data sovereignty* consists in, i.e. understandings of what its constitutive conditions are. Several different, sometimes mutually compatible claims can be distinguished.

1. **Reduction**: Some passages presented *data sovereignty* as reducible to more specific conditions and values. For example, "[M]any governments have raised concerns about national data sovereignty when government information is moved to the cloud. How can confidentiality of public information assets residing in the cloud be ensured? What if public information and IT systems are hosted abroad? Will government data of one country be caught under the authority of another jurisdiction?" (Irion, 2012: 41).
2. **Ability:** Others talk more directly about features of *data sovereignty* itself (as opposed to conditions in which it consists). For some, *data sovereignty* is something like an ability, a result of being put in a position to attain certain feats. For example, "we consider data sovereignty to be the ability of the user to have full control over his data" (Alboaie and Cosovan, 2017: 86), and "establishing data sovereignty (that is, controlling and verifying the data's geolocation) is of pivotal importance" (Esposito et al., 2016: 14).
3. **Right**: For others, *data sovereignty* is a right. For example, "data sovereignty is the right of a nation to collect and manage its own data" (Rainie et al., 2017: 5–6). In particular, "Indigenous data sovereignty is the right of Indigenous peoples and nations to govern the collection, ownership, and application of data about their peoples, lands, and resources" (Garrison et al., 2019: 506). Others do not identify data sovereignty with a right, but maintain a connection between both. For example, "[t]he rise of Indigenous data sovereignty, as an Indigenous-led movement and as a field of research, has underscored the clear rights and interests that indigenous peoples, including Māori, have in relation to Indigenous data" (Kukutai and Cormack, 2018: 145). Similarly, "[a] recommendation [of (Bigo et al., 2012)] is that existing [US–EU Safe Harbor] derogations must be disapplied for cloud due to the 'systemic risk' in regard to the potential 'loss of data sovereignty'. The report says the EU should enter into fresh negotiations with the U.S. for the recognition of a human right to privacy where Europeans are granted equal protections in U.S. courts" (Seddon and Currie, 2013: 236).

4. **Law:** *Data sovereignty* is also taken to be the outcome of legislation, often regarding the geolocation of data. For example, "Data sovereignty is the concept that information, which has been converted and stored in binary digital form, is subject to the laws of the country in which it is located" (Hippelainen et al., 2017: 645). With regards to the cloud services market, *data sovereignty* is understood to refer to "rules governing how and where certain data sets should be stored within national boundaries and outlining the rights of governments to access that data whenever they need to do so" (Courtney, 2013: 60). Others stress more defense-related aspects: "From Indonesian policy makers' perspective, the term 'data sovereignty' is not yet in use, but it refers to the national legislation on the state defence against external threats such as state-actors and non-state actors" (Nugraha et al., 2015). *Data sovereignty* concerns the extent and limitations of control over data and the cloud, both of which are prima facie subject to "competing claims of sovereign authority" (Woods, 2018: 333) from several nations and complicated by "extraterritorial effects" (ibid.: 335) of attempts to regulate the internet. This is why the shaping of *data sovereignty* must eventually be the result of "data-sovereignty litigation" (ibid.: 328).
5. **Tapping into "data wealth":** This being said, some take *data sovereignty* to contrast with legislation, e.g. data protection law: "[s]eparate to the issue of so-called data sovereignty is verification of compliance requirement, which relates data security and protection, or the rights of individuals to have control over data pertaining to them" (Courtney, 2013: 62). The promise of *data sovereignty* is that it is more innovation-friendly and flexible than rigid data protection law: "[t]he former Minister of the Economy Sigmar Gabriel thus pleaded for replacing the notion of data protection with that of data sovereignty ('Datensouveränität') on the IT summits of 2015 and 2016—a concept and guiding idea that has since been recurring in policy documents concerned with the digital economy. Gabriel also stated [. . .] that data minimization could not continue to be a guiding principle and that it would threaten the country's international competitiveness. Instead of aiming for minimization and data parsimony ('Datensparsamkeit'), a mentality of 'data wealth' was to be established, thus directly linking the proliferation of data with desirable economic outcomes" (König, 2017: 9).
6. **Relation to sovereignty tout court:** Notable differences also concern the relation between *data sovereignty* and sovereignty in general. For some, the former is a pre-condition for the latter: "[o]ne could even assert that national sovereignty is conditional upon adequate data sovereignty. If a country has no effective means of controlling public information it will become in parts dysfunctional" (Irion, 2012: 53). The same author also perceives *data sovereignty* as an extension, or an *aspect*, of sovereignty tout court: "data sovereignty poses a pertinent public policy problem for governments everywhere, because it is a crucial dimension of national sovereignty that presupposes the nation state" (Irion, 2012: 42). Specifically, *data sovereignty* instantiates national sovereignty: "At the national level, the capacity of accumulating, processing, and utilizing vast amounts of data will become a new landmark of a country's strength. The data sovereignty of a country in cyberspace will be another great power-game space besides land, sea, air, and outer spaces" (Jin et al., 2015: 60). Similarly, when Indigenous populations call for *Indigenous Data Sovereignty*, they "are reaffirming their sovereignty rights in the collection and use of Indigenous data" (Walker et al., 2018).

Given these understandings of what *data sovereignty* is, the reviewed publications outline a range of *challenges*. At least the following kinds of mutually connected issues are salient: constitutive, technical, epistemic, and legal *challenges*.

1. **Constitutive features of data:** One initial set of difficulties results from the constitutive features or the nature of data, i.e. features that are distinctive of data and set them apart from other kinds of assets and resources. For example, "[o]ne major challenge for the digital transformation of our society is the question of data sovereignty. By their very nature, unprotected data can be multiplied with very low costs and without a loss of quality" (Noll et al., 2018). Moreover, the significance of data is context-dependent: "the main criticality of data sovereignty: the definition and characterization of the area of pertinence [geographical location where data is meaningful for the application goals] is extremely mutable, and dependent on the application, and may not coincide among the one for the clouds and the one for the IoT [Internet of Things]" (Esposito et al., 2019: 4525).
2. **Technical designs:** Specific technical designs are portrayed to jeopardize data sovereignty, for example designs that aggravate complexity of data processing technologies and the ways in which they mediate between users and data subjects: "[w]ith the development of the Internet of Things, a complex CPS [cyber, physical, and social] system has emerged and is becoming a promising information infrastructure.

In the CPS system, the loss of control over user data has become a very serious challenge, making it difficult to protect privacy, boost innovation, and guarantee data sovereignty" (Yin et al., 2018). Similarly, "technical issues and logistics too often overwhelm the way data is discussed. Privileging these nuts-and-bolts kinds of considerations without first deeply considering culture, values, and Indigenous nation goals puts the cart before the horse. Therefore, this article consciously emphasizes the primary strategic issues tribes might fruitfully consider as they move toward a future vision of Indigenous data sovereignty and Indigenous nation data governance" (Rainie et al., 2017: 6). Centralized data infrastructures tend to be deemed particularly problematic, for example: "[f]requently, such a searchability is ensured by a uniform data schema and storage of the data in a central storage location, a so-called data warehouse. Nonetheless, this stands in contrast to the natural interest of the manufacturer for data sovereignty" (Hoffmann et al., 2019).

3. **Epistemic issues:** Often related to the foregoing, imperfect information complicates *data sovereignty*. For example, in cloud computing, "data sovereignty is an important problem [. . .] [M]any private and public institutions are tempted to export both their data and IT systems into the Cloud. Yet, many of them might be discouraged to do so to the extent that they cannot ensure a minimum standard of sovereignty over their own data. The difficulty to know with certainty which law applies to information stored into the Cloud creates strong legal uncertainty and raises a number of challenges that still have to be addressed by the law" (De Filippi and McCarthy, 2012: 9). Similarly, "[w]ith the increasing reliance on cloud services from software to infrastructure levels and uncertainty related to cloud security, the associated risk of data sovereignty and privacy has become higher and sometimes unclear. For example, cloud service users often do not know where the shared equipment and/or data are stored. This is an important matter because the location of data storage directly affects the applicability of local privacy legislation" (Lim et al., 2015: 92).
4. **Legal issues**: In line with the foregoing, it can be unclear which laws are applicable: "[d]espite unquestionable benefits to users, cloud computing raises several concerns about data sovereignty. Cloud providers reveal few information about geographical location and process of data and applications. As information which has been converted and stored in binary digital form is subject to the laws of the country in which it is located, it raises several concerns from a legal standpoint" (Balouek-Thomert et al., 2015).

Finally, we turn to *management strategies*, i.e. ways in which *data sovereignty* can be implemented and realized. These, too, can be stratified into constitutive, technical, epistemic, and legal proposals. Speaking to their interconnectedness, *data sovereignty* is portrayed as having multiple realizing conditions across several governance areas: "what data sovereignty should mean in practice [. . .] can of course only be reached by a mix of technical approaches, methods, and governance rules" (Jarke, 2017: 12).

1. **Promote constitutive components:** One straightforward option is to advance the constitutive components of *data sovereignty*, e.g. the rights and values with which the notion is identified or connected (see above). For example, *data sovereignty* can be realized *via* "the self-determined collection of quality data that is relevant and beneficial to indigenous communities as a key strategy of exercising indigenous sovereignty" (Fu et al., 2015: 30). Sometimes, *data sovereignty* requires attention to the *consequences* that shall be focused: "The implementation of Indigenous data sovereignty principles in the governance of Indigenous data through guidelines and practices will enhance the cultural responsiveness of biobanks and data repositories and increase Indigenous genomic research participation while protecting and honoring the rights and interests of Indigenous peoples in relation to their data" (Garrison et al., 2019: 507). Similarly, "[d]ata sovereignty means 'managing information in a way that is consistent with the laws, practices and customs of the nation-state in which it is located' [quoting (Kukutai and Taylor, 2016)]" (The First Nations Information Governance Centre, 2019: 58).
2. **Technical facilitators:** Another class of options is technical. For example, "[f]rom a technical perspective, the question remains, what are the data sovereignty protocols that need to be established" (Nugraha et al., 2015: 470), referring to encryption, national data infrastructures, and data monitoring and auditing capabilities. Similarly, others maintain that "data sovereignty should at least consist of data centre localization, national routing, national e-mail services and national backbone infrastructure" (Setiawan and Sastrosubroto, 2017: 189). Ideas like these suggest that *data sovereignty* can be *built into* IT architecture: "Since the International Data Spaces initiative offers a standardized architecture and interfaces that guarantee data sovereignty, it provides the solution to the problem just described [fear of losing control over data when sharing them]" (Otto, 2019). Some think that decentralization plays a crucial role: "data sovereignty through decentralized data storage with access via the

corresponding web services" (Doluschitz et al., 2010: 16) is key, resonating with proposals of "a method to search distributed databases, yet fully keep their owner's data sovereignty: The decentral search exploits distributed, heterogeneous, highly sensitive datasets from equally heterogeneous systems for overarching research questions. [. . .] a novel request mechanism that involves the owner with a high degree of control, who can (decentrally using their own registry or biobank systems) decide if and what to answer based on a specific project proposal" (Lablans et al., 2014).

3. **Epistemic facilitators:** Complementing the outlined epistemic issues, authors call for measures to mitigate informational shortcomings and to enhance transparency, for example when they demand "data sovereignty, which is basically that people need to know what data is collected about them and where it is being used. At the moment most people are not aware of what is being collected about them and that cannot be right. There needs to be education, awareness and approval" (Meyer, 2016: 113), and "imposing geolocation and legislation awareness policies when locating data within the cloud infrastructure" (Esposito et al., 2016: 14).
4. **Legal facilitators:** As this last suggestion indicates, legal management strategies are also proposed. For example, "some nations' data sovereignty laws require companies to keep certain types of data within the country of origin, or place significant restrictions on transmission outside the country of origin" (Vaile, 2014: 11). For businesses, "there are data sovereignty laws or industry-specific data security requirements to be reviewed" (Weeks, 2019: 9).

Agents

Table 2 presents the *agents* that co-occur with different *notions* of sovereignty. *Data sovereignty* has noticeable co-occurrences with *Indigenous population* (176 co-occurrences, c-coefficient 0.24), *user/consumer* (105, 0.14), and *private-sector organizations* (104, 0.14). The picture is more one-sided for the other notions, where *countries* is the most salient *agent* for *cyber* (169, 0.28), *internet* (91, 0.16), and *digital sovereignty* (109, 0.18).

Most often, these *agents* are portrayed as the putative sovereign, e.g. when the respective passages concern "national data sovereignty [. . .] defined here as: Government's exclusive authority and control over all virtual public assets" (Irion, 2012: 41), "Indigenous Data Sovereignty frameworks ultimately pertaining to the governance, access, collection, and use of multiple types of data by Indigenous nations and communities" (Bodkin-Andrews et al., 2019: 236), "consumers' subjective experience with regard to their own data sovereignty and thus their digital self-determination" (Krahn and Rietz, 2017: 48), or "small and medium-sized enterprises (SMEs) in the field of healthcare fear[ing] the loss of data sovereignty and information outflow" (Gembaczka et al., 2019). Occasionally, these *agents* are also portrayed as the contributors, facilitators, or difference-makers to the sovereignty of some agent or other, e.g. when commentators observe: "China's Cybersecurity Law took effect in 2017 embodying the state's commitment to 'data sovereignty'" (Kuner et al., 2019: 1); when they demand: "Data sovereignty: governments must provide and enforce laws to guarantee data independence to their citizens and companies in order to avoid particularly data access by foreign authorities" (Irain et al., 2017: 2); or when they note that "patients always have data sovereignty and further parties like care providers or researchers are only involved if desired by patients" (Dehling and Sunyaev, 2014: 89).

Contexts

The *contexts* in which the different *notions* of sovereignty appear are provided in Table 3. *Data sovereignty* shows high co-occurrence levels with the *contexts IT architecture* (215, 0.26), *legislation* (173, 0.2), and *research* (115, 0.13). In contrast, *cyber sovereignty* often co-occurs with *defense* (60, 0.21), *international relations* (41, 0.14), *legislation* (60, 0.11), and *surveillance* (30, 0.11); *digital sovereignty* with *IT architecture* (61, 0.13) and *defense* (42, 0.15); and *internet sovereignty* with *international relations* (24, 0.12) and *legislation* (42, 0.09). Again, we provide illustrations for salient *contexts* of *data sovereignty*:

1. **Legislation:** *Data sovereignty* concerns *legislation*, e.g. when "[t]he concept of 'data sovereignty'[. . .] refers to both specific data sovereignty laws limiting cross-border data transfer, as well as the more general difficulty of complying with foreign legal requirements" (Vaile, 2014: 11).
2. **IT architecture:** Data sovereignty is further debated in connection with *IT architecture*. For example, "the enhanced access to information infrastructures, content, and devices created by internet connectivity and a globalized information economy took the scale and scope of transnational surveillance to new levels. The map of its 'Worldwide SIGNINT/Defense Cryptologic Platform' leaked by Edward Snowden showed over fifty thousand computer network exploitation (CNE) implants scattered around the world, as well as 'special collection services' (SCS) in eighty locations worldwide. The exposure of these activities fueled calls for 'data sovereignty', 'technological sovereignty', and reconfigured

international cable linkages among countries both friendly and unfriendly to the United States" (Mueller, 2019: 10).

3. **Research:** In *research*, "Indigenous data sovereignty demands that research practices must be transparent about how data is stored, governed, and used" (Huria et al., 2019: 6). Moreover, "[c]onceptualizing patient information and informed consent to ensure data sovereignty is particularly challenging when data from different sources are consolidated for future research purposes" (Beier et al., 2019: 2).
4. **Societal discourse:** In a number of passages, *data sovereignty* does not merely concern data processing, but gestures towards *societal discourse and advocacy*, e.g. when the *notion* is framed as an instance of "pro-equity initiatives in industrialised economies such as the 'data sovereignty' actions of indigenous peoples and other forms of data activism" (Heeks and Shekhar, 2019: 1007). "Against this backdrop, data sovereignty requires that actors in civil society, or in cooperative economic associations, develop principles and practices that explore whether the emergent value of data should be held in common, rather than privatized; destroyed, rather than analyzed and brought to market; or stored nearby, rather than exported. [...] The objective of data sovereignty is to contest how the globalization of technology architecture (around food production or regarding other areas of social life) takes shape" (Fraser, 2018: 11). Similarly, "indigenous peoples have sovereignty over the data they have collected. This gives them the power to decide how it is used, including in holding states to account on their failure to fulfil the provisions of UNDRIP [UN Declaration on the Rights of Indigenous Peoples]" (Gilbert and Lennox, 2019: 114).

Values

As shown in Table 4, the notions *other*, *digital*, and *internet sovereignty* display high co-occurrence values only for *control and power* (0.17, 0.12, and 0.1). *Cyber sovereignty* co-occurs with *control and power* (137, 0.2) and with *security and nonmaleficence* (41, 0.11). *Data sovereignty* co-occurs with *control and power* (217, 0.2), *privacy* (108, 0.15), and further normative concepts like *deliberation and inclusion* (106, 0.14), *security and nonmaleficence* (92, 0.12), and *ownership* (80, 0.11). A closer look reveals that *data sovereignty* both advances these *values* and results from enacting them. This suggests that *data sovereignty* as used in the sample cannot be identified straightforwardly with brute force or unilateral power, but concerns more nuanced and socially embedded claims.

1. **Control and power:** *Data sovereignty* is frequently taken to involve the ability or entitlement to steer data flows and/or to govern informational resources. For example: "*Data sovereignty*: The insured person has extensive control over his/her health data to be processed in the electronic health card or the telematics infrastructure. The voluntary medical applications can be used only with the express consent of the insured person and specific access granted by him/her" (Schaar, 2010: 269). Occasionally, such characterizations already gesture towards other normative notions. For example, "the notion of data sovereignty has been established as a major guiding idea. This idea has taken on an ambivalent character in the way it has been used by the German government. While this conception of sovereignty openly aims to promote the autonomy and protection of consumers and citizens, it also locates responsibility to a greater extent in the individual. What this data sovereignty implies can be read from policy documents that are concerned with fostering a data-driven economy. It basically amounts to enabling individuals to develop relevant competences needed for having control over one's data" (König, 2017: 9). Along similar lines, others propose to speak of "data sovereignty as a burgeoning, nonlegal concept that is attractive to governments because it holds the promise of striking a balance between the progressing virtualization of information and their undiminishing demand for exclusive authority and control" (Irion, 2012: 65), and maintain that "Indigenous data sovereignty is a journey, not a destination. The journey will look different for each Native Nation [...] Tribes will use many different mechanisms on their journey [...] Central to Indigenous data sovereignty, however, is that Native Nations always remain in the driver's seat" (Carroll et al., 2019: 11). Importantly, seeking control over data can also misfire and motivate a more critical perspective: "[d]igital colonisation implies a dominant culture enforcing power and influence over minority cultures, and also takes the form of data sovereignty and disregard for digital data ownership and privacy" (Wakunuma and Masika, 2017: 2).
2. **Privacy and constraining information flows:** The *notion* further relates to *constraining* accessibility of information to others. For example, "[f]or the sake of data sovereignty of farmers, precise application maps should not be producible or reconstructable by contractors that are involved in farming activities, e.g. when applying crop protection on the field owned by a farmer. Therefore, in this scenario, the information content of tank level data should be purposefully reduced to a coarse level" (Bauer

Table 4. Heat map of the c-coefficients (co-occurrences) between notions and values. The color coding reflects the c-coefficient.

notion-value	data sov. (680)	other (314)	cyber sov. (213)	digital sov. (175)	internet sov. (110)
control and power (602)	0,2 (217)	0,17 (130)	0,2 (137)	0,12 (85)	0,1 (67)
security and non-maleficence (206)	0,12 (92)	0,08 (40)	0,11 (41)	0,09 (32)	0,02 (5)
deliberation, representation, inclusion (169)	0,14 (106)	0,08 (35)	0,03 (10)	0,06 (19)	0,02 (5)
privacy (147)	0,15 (108)	0,07 (31)	0,01 (5)	0,03 (9)	0 (1)
ownership (125)	0,11 (80)	0,09 (35)	0,01 (3)	0,03 (9)	0 (0)
transparency, epistemology (100)	0,08 (55)	0,05 (20)	0,07 (20)	0,03 (9)	0,04 (9)
effectiveness, complexity (99)	0,08 (59)	0,05 (18)	0,05 (16)	0,02 (5)	0,01 (3)
autonomy (86)	0,04 (30)	0,05 (19)	0,05 (14)	0,07 (16)	0,06 (11)
autarchy (73)	0,03 (21)	0,07 (24)	0,02 (6)	0,08 (18)	0,06 (10)
beneficence (70)	0,05 (35)	0,06 (22)	0,02 (6)	0,03 (8)	0,01 (1)
dignity, fundamental rights, identity (62)	0,03 (22)	0,09 (32)	0,01 (2)	0,03 (6)	0,02 (4)
emancipation, empowerment (48)	0,04 (27)	0,05 (17)	0 (1)	0,02 (4)	0,01 (2)
trust, reliability (48)	0,04 (28)	0,02 (8)	0,03 (7)	0,03 (6)	0,01 (1)
justice (41)	0,03 (18)	0,03 (10)	0,02 (5)	0,01 (2)	0,04 (6)
responsibility (18)	0,01 (10)	0,02 (6)	0,01 (2)	0 (0)	0 (0)
recognition, respect (15)	0,01 (7)	0,02 (5)	0,01 (2)	0 (0)	0,02 (2)

Table 5. Open-ended conceptual grid for comparing understandings of data sovereignty (one or more elements per row, no presumed vertical interdependence between elements of one column).

Agents: Who is involved?	Indigenous populations	Consumers	Countries
Contexts: What is the broader domain and/or topic?	Legislation	IT architecture	Research
Values: Which values matter?	Control and power	Privacy	Deliberation and inclusion
Descriptions: What is the primary focus?	Rights	Abilities	Legal concept
Challenges: What are the main obstacles?	Nature of data	Technical impediments	Complexity
Management strategies: How should obstacles be addressed?	Realizing rights and values	Technical designs	Focusing on consequences and effects

et al., 2019: 9). Similarly, "Indigenous data sovereignty also includes the refusal to be researched and objectified through scholarship and other data collection projects" (Dillon et al., 2019: 5).

3. **Deliberation, representation, inclusion:** More than the other *notions* in the sample, *data sovereignty* results from and requires particular modes of *deliberation* and *representation* that purposefully *include* a variety of stakeholders. For example, "data sovereignty requires that actors in civil society, or in cooperative economic associations, develop principles and practices that explore whether the emergent value of data should be held in common, rather than privatized; destroyed, rather than analyzed and brought to market; or stored nearby, rather than exported" (Fraser, 2018: 15).
4. **Ownership:** *Data sovereignty* is related to, and sometimes identified with, claims to *ownership* of data, e.g.: "*Sovereignty of medical records*: The doctor creating the medical data or the hospital storing them are not the owners of this information. Indeed, all medical information belongs to the patient. Consequently, patients have control over their data and the access to this information" (Plateaux et al., 2013). Similarly: "'data sovereignty' – personal data belongs to the user and not to the provider of an OSNs [Online Social Networks]; explicit consent of a user is necessary to sell data to third parties and a user should be able to control/track how his/her information is disseminated" (Hugl, 2011: 400). At least some ownership claims tie in with cultural and symbolic aspects: "For far too long, AIAN [American Indians, Alaska Natives] and IP [Indigenous people] have been subject to research abuses. Qualitative data, including those in big data, such as intellectual property, Indigenous knowledge, interviews, cultural expressions including songs, oral histories/stories, ceremonies, dances, and other texts, images, and recordings have been subject to and are at continued risk of exploitation, appropriation, theft, and misrepresentation. With the growing Indigenous data sovereignty movement, researchers will increasingly face not only concerns of control, ownership, and governance of research and data ownership, but also cultural and political sovereignty" (Marley, 2018: 739).
5. **Occasional lack of substantiveness:** There are instances in which *data sovereignty* is used as a buzz word without discernible meaning. For example, the claim that "[t]he main drawback of the Internet of Things is that it raises new concerns about data privacy, data sovereignty, and security" (Molina-Solana et al., 2017: 606), without any further explanation on the meaning or relation between these terms, leaves the nature of *data sovereignty* unclear. Similar remarks apply to a number of passages in which *data sovereignty* is mentioned in passing without further elaboration. For example, "[w]ith blockchain, the AI playing field can begin to level out. Independent developers can exercise fee-free ownership over their intellectual property, receive compensation for their work at a market price of their choosing, maintain data sovereignty and privacy, and transact with whom they wish in an open market" (Montes and Goertzel, 2019: 2).

Discussion

Data sovereignty is a rich, multidimensional notion with a broad range of potential connotations. Table 5 draws together the central conceptual dimensions of *data sovereignty,* illustrates possible ambiguities, provides a set of candidate meanings, and facilitates comparison between different uses even if they differ in specifics. We do not claim completeness and instead intend Table 5 to be part of an open-ended conceptual grid that can and should be complemented by further aspects.

Some connotations tend to re-appear across instances of use. *Data sovereignty* typically relates in some way to meaningful control, ownership, and other claims to data or data infrastructures. The most relevant *agents* are Indigenous populations, consumers,

and countries; the *contexts* IT architecture, legislation, and research; and the *values* control and power, deliberation and inclusion, and privacy (each of them as defined in the Appendix, Table 2). Moreover, our analysis allows for some contrastive observations of *data sovereignty* relative to other *notions*. The range of *agents* that are mentioned in connection with *data sovereignty* appears broader than for *cyber*, *digital*, and *internet* sovereignty, which primarily pertain to *countries*. While there is some overlap in the *contexts* of different *notions* of sovereignty—specifically *legislation* and *IT architecture*—*data sovereignty* has a stronger co-occurrence with *societal discourses and advocacy*, understood as the shaping of public deliberation, civil society, and collective will-formation. All notions co-occur with a broad range of *values*.

This being said, one principled issue is that authors often remain implicit or even elusive about the specifics of their understanding of *data sovereignty* and how it relates to alternative conceptions. This is precisely where the conceptual grid in Table 5 can offer potential benefits. For example, we have seen a number of instances where *challenges* or *management strategies* for *data sovereignty* have been presented without a clear elucidation of how the latter is actually understood. Recall the idea that *data sovereignty* in cloud computing is challenged by uncertainty about the physical location of data and the resulting legal uncertainty (e.g. Balouek-Thomert et al., 2015). Apart from the related claim that national data infrastructures could help to strengthen *data sovereignty*, these suggestions do not illuminate what *data sovereignty* amounts to. In such instances, there is a mismatch between the purported desirability of *data sovereignty* and the limited extent to which a positive, informative characterization is provided. We can only speculate as to why authors sometimes do not make such presuppositions explicit. There might be perceived losses of control over data that motivate and elevate the importance of speaking in terms of *data sovereignty* as a proxy for such control, while additional or alternative connotations escape attention. *Data sovereignty* might be seen as a term whose meaning is obvious and self-explanatory, and awareness of alternative meanings is lacking. It is not unusual for emerging concepts that there are different ways of making them precise. Implicit or explicit contention, controversy, and negotiation processes about what *data sovereignty* means and should mean suggest that discussants seek to leverage the *notion* towards a variety of different ends. Yet, disputants might be talking past each other or make vague policy demands if they deploy the concept without being explicit about which of the various potential connotations are intended, and how the respective claim is supported.

Such lack of clarity raises several further issues. It leaves us in the dark not only about the intentions of the authors, but also complicates reflection on the nature of *data sovereignty*. For example, some authors understand *data sovereignty* as a *right*, whereas others think that it is an *ability*. This distinction seems to mark an important difference, since it picks up on commitments regarding whether *data sovereignty* is something that is already there or possessed and thus motivates certain demands, or whether it is more like a *telos*, something towards which we should aim. Of course, these options are not strict alternatives, and there is room for mutual compatibility. For example, someone can in principle have a right to be *equipped with* certain *abilities*. And it is also possible to have the ability to claim, articulate, and insist on the enforcement of certain rights. Yet, despite not being strict alternatives, the outlined options indicate different emphases on the order of priority between *data sovereignty* and the concepts figuring in *descriptions* of it.

As another example, consider the relation between *data sovereignty* and autonomy. On the face of it, there are certain parallels between both concepts. They appear to involve a particular kind of freedom from external interferences, and positive freedom to proceed as one pleases. However, if there is a connection or analogy between sovereignty and autonomy, very few passages make it explicit. Consider at least two questions one might have. First, suppose that being in control of the flow of one's data is sufficient for *data sovereignty*, and that I exercise control over my data by proceeding to share it with an internet service provider. Have I acted autonomously? It depends. For example, Dworkin (1976) understands autonomy as freedom from external constraints plus authenticity, i.e. affirmation of a choice based on higher-order preferences. In another influential definition, Beauchamp and Childress (Beauchamp and Childress, 2013: 104–105) require intentionality, i.e. correspondence to the agent's conception of the act, understanding of what one is about to do, and lack of controlling influences that determine her action. Now, note that mere control does not by itself satisfy conditions such as authenticity or understanding. This allows for at least two interpretations. *Either* autonomy might be taken to be a stronger, more demanding notion than *data sovereignty, or* the kind of control that is necessary for *data sovereignty* is not merely control, but a particularly meaningful kind of control, e.g. one with epistemic and social presuppositions, such as awareness of the potential scope of data processing applications and entitlements to control resulting from recognition of one's fundamental rights. Our results provide resources to pursue either of these avenues. There can clearly be overlap between *data sovereignty* and autonomy, but depending on how

one makes either precise, they can also significantly come apart.

Such open questions about the nature of *data sovereignty* are entangled with a set of more practical issues. To begin with, they complicate the assessment of concrete proposals. For example, if descriptively, *data sovereignty* is understood as a *right*, then *management strategies* focusing on mere technical designs alone will probably not be enough to advance it, fall short of the position's own standards, and be discredited as technological solutionism (Mozorov, 2013): a putative technical fix to a problem that demands much more. Along similar lines, if *data sovereignty* is understood as an issue in the *context* of legislation that entitles subjects to have their data stored in certain geolocations, then presumably additional steps are required to align with the rich set of connotations of Indigenous *data sovereignty* concerning culture, inclusion, integrity, and identity.

Another practical issue arising from lack of clarity about intended meaning and/or the nature of *data sovereignty* is that it obscures the significance of negotiation processes. One reason for caution with taking claims to *data sovereignty* at face value is that several *agents* can instantiate *data sovereignty*, or articulate a reasonable claim to *data sovereignty*. While the surveyed publications occasionally mention that there can be incompatibilities and tradeoffs, e.g. between the *data sovereignty* of individuals and the *data sovereignty* of a population, society, or country, many authors seem to abstract from or even neglect this complication. When expounding and assessing positions, one crucial question becomes: *who* amongst these prima facie data sovereigns shall take precedence? And when we demand that we need to safeguard, comply with, or respect *data sovereignty*, which agent is the focal point of this normative claim?

Finally, lack of clarity about the nature of *data sovereignty* complicates the pragmatic question of what it takes to attain it and which concrete mechanisms for implementation are needed. Related to this issue, we might seek further clarification on who is responsible to ensure *data sovereignty*. For each of the *agents* addressed, what would they have to do in order to foster it? Is it primarily a responsibility of the state to issue certain laws? Does the private sector carry responsibility through self-commitment? Is it the users who need to educate themselves and adapt consumption behavior accordingly? Who adjudicates in cases of *conflicts* between claims from different putative data sovereigns, e.g. when national *data sovereignty* undercuts individual citizen *data sovereignty*, or vice versa? References to *data sovereignty* alone can seem devoid of substantive content on how to handle such conflicts. Again, authors occasionally touch upon these issues, but systematic explorations are surprisingly rare, and many times these questions remain unaddressed.

In light of these issues, Table 5 thus addresses a need by mapping different understandings and connotations, and could pave the way to enhanced conceptual and argumentative clarity. While we do not intend to evaluate diverging understandings of *data sovereignty* or to argue for certain practical implications in the political sphere on the basis of our review, our results suggest that care should be taken to disambiguate between the various candidate meanings, and to spell out what exactly *data sovereignty* is supposed to involve and require in a given context. Besides offering a non-exhaustive frame that serves an urgent need to systematize mentions of *data sovereignty*, Table 5 can facilitate subsequent assessment, negotiation, and implementation processes.

A further critical point for discussion and exploration is that parts of the debate are oddly compartmentalized. In numerous publications, commentators on Indigenous *data sovereignty* unfold a rich and fascinating *notion* that stands out for a number of reasons. First, it ties *data sovereignty* to fundamental features of the *agent*, such as her culture and identity, and thus marks a particularly intimate relationship between exercising control over data and the integrity of the data sovereign. Second, Indigenous *data sovereignty* is portrayed as fully continuous with the established sovereignty of the respective Indigenous Nation. Third, Indigenous *data sovereignty* involves control over data, but also requires involvement in deliberation on data governance and societal discourses on how to harness data. Fourth, proponents of the *notion* harness its emancipatory aspects and leverage it towards criticizing asymmetries of power, established structures, and historically manifested injustices. All these points are occasionally touched upon or implicitly alluded to by other discussions of *data sovereignty*, too, but Indigenous *data sovereignty* raises them consistently. Surprisingly, however, despite these rich and innovative aspects of Indigenous *data sovereignty,* other discussants acknowledge it only rarely. For example, the reviewed publications from the German discourse do not mention it even in passing.

As outlined in the "Introduction and Methods" section, our study has limitations. First, its scope is limited insofar as it focuses on academic writing. It would be interesting to extend similar analyses of *data sovereignty* to other fields, such as journalism and social media content. Second, no quality assessment of passages mentioning *data sovereignty* was made, although our focus on academic writing means that peer-review processes have played some role in the returned search results that were screened for inclusion and exclusion.

Third, our study takes no stance on how to advance *data sovereignty* in practice. Our results describe strategies mentioned in the literature, but we do not assess the appropriateness of these strategies. Fourth, we carved up the data inductively and without applying an antecedently set theory to the material. While we find this approach appealing and are convinced of the soundness of the inductive scheme, we do not rule out that there might be other, similarly acceptable ways to make sense of the data. Fifth, the scheme refers to concepts that are themselves broad and/or not used uniformly across the literature. To contain this limitation as far as possible, we deployed general working definitions (described under "Methods" section and Table 2 in the Supplemental Appendix) that we think are reasonably general and precise.

Conclusion

We have reviewed how the notion of *data sovereignty* is used in academic writing. The *notion* turned out to exhibit a variety of different candidate meanings, and we have presented a conceptual grid to systematize them. The candidate meanings tend to relate in some way to meaningful control, ownership, and other claims in data. *Data sovereignty* can apply to a range of *agents* across the spectrum from individual consumers to entire societies and countries, sometimes yielding conflicting claims to *data sovereignty* across these levels. It primarily occurs in the *context* of debates around the design of IT architecture and/or laws applicable to data processing, but a number of other *contexts* as well. It tends to address a nuanced mixture of *values*: typically, it concerns *control and power* over data, yet the kind of power in question is not brute an arbitrary power, but often ties in with considerations related to inclusive deliberation and fundamental rights of data subjects. Finally, distinguishing more sharply and explicitly between *descriptions* of *data sovereignty*, *challenges* to *data sovereignty*, and *management strategies* to overcome them could ameliorate discourses and negotiation processes surrounding the governance of the digital.

Acknowledgements

The authors are grateful to Serena Bischoff, Simone Donner, Sebastian Hummel, David Samhammer, and to three anonymous reviewers for their helpful comments and suggestions.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The authors are grateful for funding from the German Federal Ministry of Health (Project DABIGO; ZMV/1 – 2517 FSB 013) and the German Ministry of Education and Research (Project CwiC; 01GP1905B). The funders played no role in planning, designing, and conducting the study.

ORCID iD

Patrik Hummel https://orcid.org/0000-0001-9668-0810

Supplemental material

Supplemental material for this article is available online.

References

Alboaie S and Cosovan D (2017) Private data system enabling self-sovereign storage managed by executable choreographies. *Lecture Notes in Computer Science LNCS* 10320: 83–98.

Baezner M and Robin P (2018) Cyber sovereignty and data sovereignty. *CSS Cyber Defense Project*.

Balouek-Thomert D, Caron E, Gallard P, et al. (2015) Nu@ge: Towards a solidary and responsible cloud computing service. In: *International conference on cloud technologies and applications*, Marrakech, Morocco, 2–4 June 2015.

Bauer J, Helmke R, Bothe A, et al. (2019) CAN't track us: Adaptable privacy for ISOBUS controller area networks. *Computer Standards & Interfaces* 66: 103344.

Beauchamp TL and Childress JF (2013) *Principles of Biomedical Ethics*. Oxford and New York: Oxford University Press.

Beier K, Schweda M and Schicktanz S (2019) Taking patient involvement seriously: A critical ethical analysis of participatory approaches in data-intensive medical research. *BMC Medical Informatics and Decision Making* 19(1): 90–10.

Bigo D, Boulet G, Bowden C, et al. (2012) *Fighting Cyber Crime and Protecting Privacy in the Cloud*. Brussels: European Parliament, Directorate General for Internal Policies.

Bodkin-Andrews G, Page S and Trudgett M (2019) Working towards accountability in embedding indigenous studies: Evidence from an indigenous graduate attribute evaluation instrument. *Australian Journal of Education* 63(2): 232–260.

Carroll SR, Rodriguez-Lonebear D and Martinez A (2019) Indigenous data governance: Strategies from United States native nations. *Data Science Journal*. DOI: 10.5334/dsj-2019-031.

Courtney M (2013) Regulating the cloud crowd. *Engineering & Technology* 8(4): 60–63.

Couture S and Toupin S (2019) What does the notion of "sovereignty" mean when referring to the digital? *New Media & Society* 21(10): 2305–2322.

De Filippi P and McCarthy S (2012) Cloud computing: Centralization and data sovereignty. *European Journal of Law and Technology* 3(2): 1–18.

Dehling T and Sunyaev A (2014) Secure provision of patient-centered health information technology services in public networks—leveraging security and privacy features provided by the German nationwide health information technology infrastructure. *Electronic Markets* 24(2): 89–99.

Dillon L, Lave R, Mansfield B, et al. (2019) Situating data in a Trumpian era: The environmental data and governance initiative. *Annals of the American Association of Geographers* 109(2): 545–555.

Doluschitz R, Engler B and Hoffmann C (2010) Quality assurance and traceability of foods of animal origin: Major findings from the research project IT FoodTrace. *Journal für Verbraucherschutz und Lebensmittelsicherheit* 5(1): 11–19.

Dworkin G (1976) Autonomy and behavior control. *The Hastings Center Report* 6(1): 23–28.

Esposito C, Castiglione A and Choo KKR (2016) Encryption-Based Solution for Data Sovereignty in Federated Clouds. *IEEE Cloud Computing* 3(1): 12–17.

Esposito C, Castiglione A, Frattini F, et al. (2019) On data sovereignty in cloud-based computation offloading for smart cities applications. *IEEE Internet of Things Journal* 6(3): 4521–4535.

Floridi L (2020) The fight for digital sovereignty: What it is, and why it matters, especially for the EU. *Philosophy & Technology* 33: 369–378.

Fraser A (2018) Land grab/data grab: Precision agriculture and its new horizons. *Journal of Peasant Studies*. DOI: 10.1080/03066150.2017.1415887. 1-20.

Friese S (2019) *ATLAS.ti 8 Mac - User Manual*. Berlin: ATLAS.ti Scientific Software Development GmbH.

Fu M, Exeter DJ and Anderson A (2015) "So, is that your 'relative' or mine?" a political-ecological critique of census-based area deprivation indices. *Social Science & Medicine (1982)* 142: 27–36.

Garrison NA, Hudson M, Ballantyne LL, et al. (2019) Genomic research through an indigenous lens: Understanding the expectations. *Annual Review of Genomics and Human Genetics* 20: 495.

Gembaczka P, Heidemann B, Bennertz B, et al. (2019) Combination of sensor-embedded and secure server-distributed artificial intelligence for healthcare applications. *Current Directions in Biomedical Engineering* 5(1): 29–32.

German Ethics Council (2017) *Big Data and Health — Data Sovereignty as the Shaping of Informational Freedom (Executive Summary & Recommendations)*. Berlin: German Ethics Council.

Gilbert J and Lennox C (2019) Towards new development paradigms: The United Nations Declaration on the Rights of Indigenous Peoples as a tool to support self-determined development. *The International Journal of Human Rights* 23(1–2): 104–124.

Heeks R and Shekhar S (2019) Datafication, development and marginalised urban communities: An applied data justice framework. *Information, Communication & Society* 22(7): 992–1011.

Hippelainen L, Oliver I and Lal S (2017) Towards dependably detecting geolocation of cloud servers. In: Yan Z, et al. (eds) *Network and System Security*. Cham: Springer, pp.643–656.

Hoffmann A, Wagner A, Huyeng T, et al. (2019) Distributed manufacturer services to provide product data on the web. In: *CEUR workshop proceedings, EG-ICE 2019 Workshop on Intelligent Computing in Engineering*, Leuven, Belgium, 30 June–3 July 2019.

Hugl U (2011) Reviewing person's value of privacy of online social networking. *Internet Research* 21(4): 384–407.

Hummel P, Braun M, Augsberg S, et al. (2018) Sovereignty and data sharing. *ITU Journal: ICT Discoveries* 2.

Huria T, Palmer SC, Pitama S, et al. (2019) Consolidated criteria for strengthening reporting of health research involving indigenous peoples: the CONSIDER statement. *BMC Medical Research Methodology* 19(1): 173–179.

Irain M, Jorda J and Mammeri Z (2017) Landmark-based data location verification in the cloud: Review of approaches and challenges. *Journal of Cloud Computing* 6(1): 1–20.

Irion K (2012) Government cloud computing and national data sovereignty. *Policy & Internet* 4(3–4): 40–71.

Jarke M (2017) Data spaces: Combining goal-driven and data-driven approaches in community decision and negotiation support. In: Schoop M and Kilgour DM (eds) *Lecture Notes in Business Information Processing* 293. Cham: Springer, pp.3–14.

Jin X, Wah BW, Cheng X, et al. (2015) Significance and challenges of big data research. *Big Data Research* 2(2): 59–64.

König PD (2017) The place of conditionality and individual responsibility in a "data-driven economy". *Big Data & Society* 4(2): 205395171774241.

Krahn B and Rietz C (2017) Consumers' digital self-determination: Everything under control? In: Linnhoff-Popien C, et al. (eds) *Digital Marketplaces Unleashed*. Berlin: Springer, pp.45–55.

Krasner SD (1988) Sovereignty. An institutional perspective. *Comparative Political Studies* 21(1): 66–94.

Kukutai T and Cormack D (2018) Census 2018 and implications for Māori. *New Zealand Population Review* 44: 131–151.

Kukutai T and Taylor J (2016) *Indigenous Data Sovereignty. Toward an Agenda*. Canberra: Australian National University Press.

Kuner C, Cate FH, Lynskey O, et al. (2019) Introducing the global data privacy prize. *International Data Privacy Law* 9(1): 1–1.

Lablans M, Kadioglu D, Muscholl M, et al. (2014) Preserving the owner's autonomy in networks of patient registries and biobanks. *Orphanet Journal of Rare Diseases* 9(9042): P3.

Lim N, Grönlund Å, Andersson A, et al. (2015) Cloud computing: The beliefs and perceptions of Swedish school principals. *Computers & Education* 84: 90–100.

Marley TL (2018) Indigenous data sovereignty: University Institutional Review Board policies and guidelines and research with American Indian and Alaska native communities. *American Behavioral Scientist*. DOI: 10.1177/0002764218799130. 276421879913.

Meyer H (2016) On technology innovations, digitalisation and social security in the 21st century: Interview with Henning Meyer. *Communication Today* 7(2): 108–115.

Moher D, Liberati A, Tetzlaff J, et al. (2009a) Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine* 6(7): e1000097.

Molina-Solana M, Ros M, Ruiz MD, et al. (2017) Data science for building energy management: A review. *Renewable and Sustainable Energy Reviews* 70: 598–609.

Montes GA and Goertzel B (2019) Distributed, decentralized, and democratized artificial intelligence. *Technological Forecasting and Social Change* 141: 354–358.

Mozorov E (2013) *To Save Everything, Click Here: The Folly of Technological Solutionism*. New York: PublicAffairs.

Mueller ML (2019) Against sovereignty in cyberspace. *International Studies Review*. DOI: 10.1093/isr/viz044.

Noll F, Lüke W, Wappler M, et al. (2018) Industrial companies in innovation ecosystems. In: *2018 IEEE international conference on engineering, technology and innovation (ICE/ITMC)*, Stuttgart, Germany, 17–20 June 2018, pp.1–5.

Nugraha Y, Kautsarina K and Sastrosubroto AS (2015) Towards data sovereignty in cyberspace. In: *2015 3rd international conference on information and communication technology (ICoICT)*, Nusa Dua, Indonesia, 27–29 May 2015, pp.465–471.

Otto B (2019) Interview with Reinhold Achatz on "data sovereignty and data ecosystems". *Business & Information Systems Engineering* 61(5): 635.

Plateaux A, Lacharme P, Rosenberger C, et al. (2013) A contactless e-health information system with privacy. In: *2013 9th international wireless communications and mobile computing conference (IWCMC)*, Sardinia, Italy, 1–5 July 2013, pp. 1660–1665.

Rainie SC, Schultz JL, Briggs E, et al. (2017) Data as a strategic resource: Self-determination, governance, and the data challenge for indigenous nations in the United States. *International Indigenous Policy Journal* 8(2).

Schaar P (2010) Privacy by design. *Identity in the Information Society* 3(2): 267–274.

Seddon JJM and Currie WL (2013) Cloud computing and trans-border health data: Unpacking U.S. and EU healthcare regulation and compliance. *Health Policy and Technology* 2(4): 229–241.

Setiawan AB and Sastrosubroto AS (2017) Strengthening the security of critical data in cyberspace, a policy review. In: *2016 international conference on computer, control, informatics and its applications (IC3INA)*, Tangerang, Indonesia, 3–5 October 2016, pp.185–190.

Strech D and Sofaer N (2012) How to write a systematic review of reasons. *Journal of Medical Ethics* 38(2): 121–126.

The First Nations Information Governance Centre (2019) First nations data sovereignty in Canada. *Statistical Journal of the IAOS* 35(1): 47–69.

Tranfield D, Denyer D and Smart P (2003) Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *British Journal of Management* 14(3): 207–222.

Vaile D (2014) The Cloud and data sovereignty after Snowden. *Australian Journal of Telecommunications and the Digital Economy* 2(1): 31–31.

Wakunuma K and Masika R (2017) Cloud computing, capabilities and intercultural ethics: Implications for Africa. *Telecommunications Policy* 41(7–8): 695–707.

Walker J, Healy B, Healy C, et al. (2018) Perspectives on linkage involving indigenous data. *International Journal of Population Data Science* 3(4): 1–2.

Weeks R (2019) How to implement a robust BCDR plan. *Computer Fraud & Security* 2019(7): 8–11.

Winandy M (2011) A note on the security in the card management system of the German E-Health Card. In: *Electronic healthcare – third international conference, eHealth 2010*, Casablanca, Morocco, 13–15 December 2010, pp. 196–203.

Woods AK (2018) Litigating data sovereignty. *Yale Law Journal* 128(2): 328–406.

Yin H, Guo D, Wang K, et al. (2018) Hyperconnected network: A decentralized trusted computing and networking paradigm. *IEEE Network* 32(1): 112–117.