Unskewing the Data Value Chain

The Histories, Practices and Policies of Community Data Governance in the Global South

Aditya Singh & Divij Joshi

IT for Change

Acknowledgments

Research & Writing: Aditya Singh and Divij Joshi

Editorial Support: Deepti Bharthur

Copy Editing & Proofreading: Sohel Sarkar and Intifada P. Basheer

Layout: Jojy Philip

Contents

The Histories, Practices and Policies of Community Data Governance in the Global South

ADITYA SINGH AND DIVIJ JOSHI

1. Introduction: The Data Economy and North-South Divides

Data and data analytics technologies like artificial intelligence (AI) are transforming digital economies. The production of knowledge and intelligence through data has become so important to value chains in the digital economy, that the phrase "data is the new oil" has become commonplace (The Economist, 2017). The metaphor invokes the centrality of data and information technologies as an economic infrastructure for a transformed digital economy and underscores the eagerness of businesses and governments looking to harness the potential of data for economic development.

However, the nature of the emerging globalized digital economy is characterized by inequities. The sources of these inequities are rooted within the peculiar norms, institutional forms, and infrastructures that constitute the digital economy, which prioritize the 'free market' of data and enable the monopolization of economic and political power by Big Tech (Cohen, 2019). The terrain of inequitable development of the digital economy is also not uniform across geographies. The forms that the digital economy takes in the Global South, particularly among post-colonial nations, continue to be influenced by historical-developmental inequities. These inequities are evidenced by international trade regimes like the Regional Comprehensive Economic Partnership (RCEP), which inhibit sovereign control over data and algorithms, or by global tax regimes that have allowed Big Tech to evade domestic tax regimes in the Global South (Vipra, 2019).

Recent policy responses and positions from the Global South have begun to acknowledge these structural inequities in the political economy of the globalized digital economy, and the necessity to center data and information as a key element of policymaking and governance. While data governance policy has historically centered around considerations of information as property (as in intellectual property or copyright) or individual data protection regimes, emerging trends in the field have recognized the limitations of these conceptions. These new trends are informed by alternative conceptions of data – including information as a sovereign resource, or as a matter of governance by communities – which must be relationally and contextually informed.

However, in the post-colonial economies of the Global South, competing assertions of ownership and sovereignty over natural resources like oil, have led to their own inequitable political economies and can risk becoming a violent enterprise. This is evidenced by the displacement of indigenous populations from their lands, done in the name of economic development (Kovaks & Ranganathan, 2020). Furthermore, in responding to inequitable information systems, regulations from the Global South must not only confront the historical patterns of exploitation across geographies, but also the inherently extractive nature of contemporary information systems and data governance models.

> "While data governance policy has historically centered around considerations of information as property or individual data protection regimes, emerging trends in the field have recognized the limitations of these conceptions. These new trends are informed by alternative conceptions of data – including information as a sovereign resource, or as a matter of governance by communities – which must be relationally and contextually informed."

In this paper, we argue that there is substantial scope for data governance policies for the Global South to draw from theories and practices informed by the idea that data can be a subject of decentralized, community-centric governance. First, we highlight the importance of recognizing communities and groups as agents with interests and rights in various forms of data, and how this challenges many of the assumptions upon which contemporary data policy and globalized information systems are built. Second, we explore how the idea of community data governance is informed by notions of knowledge commons, and the importance of the theoretical framework of commons governance to information and knowledge economies. Third, we identify three distinct theories and practices of community data governance, which have responded to concerns about power, information, and knowledge in the post-colonial context, and identify how these theories and practices can inform data governance policy in the Global South.

2. Community-centric Data Governance

Recognizing data as relational and focusing on groups and communities to whom it relates, challenges many of the ontological and epistemic claims on which current responses to inequitable data governance practices rest – such as the centrality of ownership rights and individualistic conceptions of data protection. Various technical and legal measures have been proposed in order to provide alternative imaginations of data governance. Particularly, of late, there has been increasing attention to the shortcomings of data governance models that center the individual (Lehtiniemi & Kortesniemi, 2017). For instance, Delacroix and Lawrence's articulation of "data trusts", an organizational mechanism for ensuring trustworthy governance of individual or community data, is explicitly premised upon the limited capacity and power individuals possess in the data economy, relative to the data controllers they are expected to

transact with (Delacroix & Lawrence, 2019).

Such proposals are in line with broader critiques that note the inability of current data protection regimes, premised on individual identifiability and individual harm, to fully address the impact of contemporary data practices. This has brought attention to 'group privacy' and collective conceptions of privacy as a means of conceiving protection from harm. Contemporary data practices are no longer premised on the identifiability of individuals, but on deriving insights from groups of people with shared characteristics, which may then be used to target populations with similar features (Taylor, 2017). The objectives may range from online targeted advertising to more egregious forms of social sorting. The implication is also that certain groups may be more vulnerable to data-driven harms. Exacerbating the vulnerabilities of socially marginalized groups is a harm that is normatively distinct from concerns related to individual autonomy that motivate much of current privacy and data protection regulation (Viljoen, 2020).

> "Recognizing data as relational and focusing on groups and communities to whom it relates, challenges many of the ontological and epistemic claims on which current responses to inequitable data governance practices rest – such as the centrality of ownership rights and individualistic conceptions of data protection."

As Viljoen notes, current legal regimes focus exclusively on the vertical relationship between data controllers and data subjects, but not on the downstream, horizontal relationships that are created by controllers between data subjects in the data economy. However, creating these horizontal relationships, by generating population-level insights from groups, to target populations of similar characteristics is the key economic driver of contemporary data practices (Viljoen, 2020). This implies that the scope of analysis of data governance must address these relationships more fully. This requires recognition that not just individuals, but communities, algorithmic or self-constituted, may also be a relevant unit of analysis.

The task of data governance, then, is not limited to extending individual control over their vertical relationships with data, but to creating institutional responses to data practices that allow stakeholders to negotiate interests and preferences. This will create space for communities to be a focal point for evaluating legitimate data practices, including socially beneficial data production and stewardship. Strong protections only focused on individual control risk foreclosing the potential benefits from datafication, those with appropriate legitimacy, transparency, and social license.

3. Information Commons and Data Governance

The management of community knowledge and information resources can draw from the theory and practice of the commons, which have long represented and studied community-led stewardship of shared resource systems. The analytical resources employed by these systems demonstrate how communities may assert their sovereignty

over a resource, and navigate contestations around boundaries, values, and use. Recent literature has already begun to address issues of data governance using the framework of the commons.

The commons represent a tradition of theory and practice examining how communities can sustainably manage shared resources without assigning private property rights or resorting to centralized government management. The concept was popularized by Ostrom in her critique of Garrett Hardin's 'Tragedy of the Commons' (Ostrom, 1990).

Ostrom wrote her seminal critique of the 'Tragedy of the Commons' by demonstrating thousands of instances where communities had sustainably managed shared resources without resorting to private property rights or to centralized management from governments. Examples ranged from forest communities in Nepal, to pastures in the Swiss Alps and irrigation systems in the Valencia region. Most notably, she identified "design principles" which were largely present in all instances where common resources had been sustainably and equitably governed.

> "The task of data governance, then, is not limited to extending individual control over their vertical relationships with data, but to creating institutional responses to data practices that allow stakeholders to negotiate interests and preferences. This will create space for communities to be a focal point for evaluating legitimate data practices, including socially beneficial data production and stewardship."

Ostrom's analysis was originally restricted to what she described as "common-pool resources" – goods characterized by high subtractability[1] and low excludability[2] (these were typically natural resource commons such as community grazing lands, fisheries, etc.). However, the analytical lens of the commons has been subsequently extended to other resources that need not exhibit these characteristics. It is argued that the ability for a resource to be held in common is not a property inherent to a resource, but a function of the institutional arrangements that support shared use (De Angelis, 2017). Bollier emphasizes that "it is not the good that is excludable or not, it's people who are being excluded or not" (Bollier & Hefrich, 2019). The commons, thus presents a bottom-up framework of governance that allows a community to match rules with local contexts, by granting the community the autonomy to experiment with diverse sets of rules. Facilitating democratic decision-making allows the community to prevent inequitable outcomes (Bollier & Hefrich, 2019). Commons may thus arise from the intention and action of a community to manage a resource collectively, with a view towards fair and sustainable access, use, and governance.

This extension of the commons is exemplified by the more recent study of 'knowledge commons', which deals with intangible resources like knowledge, data, and information. In particular, scholars found that the commons could help conceptualize new dilemmas

[1] When the use of a good by one person reduces the availability of that good for others.

[2] Implies the ease with which individuals may be excluded from the use of a good.

being observed with the rise of distributed digital information (Hess & Ostrom, 2007). A vast array of knowledge commons arrangements has since been studied, including indigenous knowledge, scientific data, open-source software, and knowledge repositories like Wikipedia (Frischmann et al., 2014).

The centrality of data to the modern economy has prompted discussion on commons-based governance frameworks as a model that could enable greater sharing of data in ways that can align with community values. For instance, Bambauer describes data being enclosed in private datasets as a 'Tragedy of the Data Commons' (Bambauer, 2012).

While some conceptions of 'data commons' focus solely on their technical attributes (Grossman et al., 2016), recent literature has brought attention to commons as a model for data sharing within defined communities. Beckwith et al. argue that in the context of smart cities, the commons may be a more appropriate institutional framework for data access than open data (Beckwith et al., 2019). Open data doesn't acknowledge the existence of real communities, where there may be greater comfort in sharing within an institutional framework that protects from outsiders and potential harm. Baarbé et al. also propose data commons in agriculture, aimed towards achieving food security (Baarbé et al., 2017). Furthermore, they note that data commons could apply beyond agriculture to other domains as well, emphasizing the salience of commons as providing a framework that enables engaging all stakeholders.

As highlighted by the emerging literature on group privacy and the collective dimensions of privacy, the highly relational nature of data calls into question the centrality of the individual in current approaches to data governance. Evans presents the commons as institutional arrangements for managing or governing the production and use of such a resource, while addressing dilemmas that impede sustainable sharing and stewardship (Evans, 2016). She notes that in the context of health data, there are strong incentives for greater sharing of non-de-identified, deeply descriptive data, but its relational nature implies that individual-centric consent and control models are unworkable. In the context of data, people have the ability to make decisions that affect others and their data. This indicates the relevance of commons-based governance. In comparison, data governance stemming from autonomy-based bioethics disempowers the people it seeks to protect when it precludes collective action on matters of common interest. Evans notes that legal systems employ informed consent for decisions that affect the individual, but not for matters of public safety and welfare. Bioethical principles that grant individuals veto rights over the use of data "blur the line between individual autonomy and narcissism" (Evans, 2016).

Ho and Chuang, similarly, note the limitations of individual-centric data governance regimes and propose a commons-based sharing framework (Ho & Chuang, 2019). They emphasize that trust within the commons can facilitate greater sharing between commoners as opposed to entities outside communal bounds. They argue that initiatives like Wikipedia, OpenStreetMaps, and Social.Coop demonstrate that data

and information can be aggregated, shared, and managed by a community of peers. In particular, computational methods (such as security, multiparty computation, open audit e-voting, and user-centric online services) could help achieve the context-dependent confidentiality and auditability requirements that may be necessary for data commons.

Prainsack argues that Ostrom's design principles, as originally proposed, may not be directly applicable to data as a resource (Prainsack, 2019). Given the specific nature of their materiality (for instance, the fact that they're easily replicated, and possibly exist simultaneously at various locations), they cannot be governed as common-pool resources based on Ostrom's design principles. When organizing data and information as commons, Prainsack instead emphasizes the importance of inclusion and exclusion, particularly when the focus is on counteracting power asymmetries. In particular, she argues that all information commons should have mechanisms that allow scrutiny over practices and requirements that have bearing on undue exclusion from the use and governance of the resource. She presents a restatement of the design principles that foregrounds such an analysis of exclusion.

The commons can, therefore, order thinking around how data governance institutions' response to community needs and interests may be imagined. However, the discourse on redistributing power in the digital economy must additionally consider the specific contexts, histories, and intellectual traditions of data governance in the Global South, and not simply consolidate and reproduce patterns of North-led norm formation.

4. Principles and Practices of Community Knowledge Management

If data policy for the Global South is to focus on empowering communities, rather than reproducing the racist, colonial, and extractive infrastructures and practices of contemporary information capitalism, we argue that such policy should be informed by the theories and practices of information and knowledge management, which have been developed by and for communities. Furthermore, such policy should derive from methodologies and interventions, which recognize the inherently political nature of information and knowledge systems. In particular, we identify three distinct but interrelated practices of community data governance, which draw from distinct theoretical and disciplinary lenses, but converge on ideas of recognizing communities as agents in the governance of information and information systems.

4.1 Community Archives

Archives as spaces for the collection, presentation, and retrieval of current and historically valuable information, present distinct theories and practices for the governance of information and knowledge systems. Archives, generally, represent the collection of 'traces' of information and knowledge, which are considered to have lasting value in particular contexts. In doing so, they draw from specific principles

such as provenance, order, custody, value, authenticity, and standardized systems of arrangement, classification, and description (Harris, 2002). The practices of archival information management have long considered questions relevant to contemporary digital information systems, including questions of privacy and data protection, attribution, and valuation of archival records. Correspondingly, the development of digital information systems draws from traditions and practices developed in the context of physical archiving. As Agostinho et al. note, in the context of Big Data and the contemporary information economy, "Current practices of data production, collection, distribution, and consumption both build upon and draw from the cultural history of the archive, as well as raising pertinent new questions that exceed the horizon of physical archives" (Agostinho et al., 2019).

In particular, critiques of information systems that draw from cultural theories of the archives recognize that the nature and forms of an archive function not only as a mechanism for collecting, storing, and retrieving information or knowledge, but as a construct and context that defines and produces knowledge through decisions about what gets archived, and how preservation and access mechanisms privilege certain forms of power and control over information (Agostinho et al., 2019). In doing so, they necessarily act as mediators of power and knowledge, and often in ways that reproduce historical inequities. Archival theory and practice have sought to recognize and respond to these critiques in numerous ways.

> "The centrality of data to the modern economy has prompted discussion on commons-based governance frameworks as a model that could enable greater sharing of data in ways that can align with community values."

For the purpose of this paper, we draw from the theories and practices relating to 'community archives'. Community archiving escapes strict definition and broadly encompasses archival institutions or practices which are "grassroots, community-owned-and-controlled initiatives that collect, describe, and make accessible materials of the community's own choosing on its own terms" (Poole, 2020). As such, community archives serve as an important organizational and institutional form that considers the governance of information and knowledge explicitly through the identification of group or community interests within such information.

Community archiving makes practices of participation and engagement of communities who have interests in archival records and information more explicit, and encompasses a variety of mechanisms to involve communities in practices of archival appraisal, custody, preservation, valuation, and access.

Community archives respond to the role played by legal and physical ownership and custody of records within archives, as a mode through which power is negotiated between the creators of a record or information, the archivists or the repository, and the users who can access the record. Many efforts towards community archiving draw from

the concept of 'post-custodial' archiving, which explicitly endorses that the function and role of an archivist should move beyond maintaining ownership and custody of records, towards facilitating stewardship of the records on behalf of the community which has interests in the information organized within the archive. According to Wurl, "A stewardship ethos encompasses a very different set of relationships between stakeholders and materials. It is characterized by partnership and continuity of association between repository and originator. In a stewardship approach, archival material is viewed less as property and more as a cultural asset, jointly held and invested in by the archive and the community of origin" (Wurl, 2005). In the turn towards post-custodial stewardship, archival practice attempts to renegotiate the dynamics of power and control over information and knowledge resources between the creators of a record, the archivists, and the users of the archive.

Community archiving reconfigures information categorization, classification, and management practices in ways that enable community participation. Community archives in the Global South have embraced post-custodial approaches towards archiving, where professional archivists privilege mechanisms for building trust and enabling access among a community's records, over bureaucratic requirements of ownership and physical custody (Halim, 2018).

This also reflects in methodologies for appraisal of records which are decentralized and draw from community values and volunteer experience. Moreover, information labeling and categorization systems are reconfigured in ways that allow for communities to understand, locate, and evaluate records towards their own ends, rather than towards the ends of other institutions or external research (Montenegro, 2019).

4.2 Indigenous Data Sovereignty

Indigenous data refers to any data, in any format, that relates to indigenous peoples, lands, resources, communities, lifeways, and cultures (Raine et al., 2019). As noted by Walter et al., such data define how the state "sees" (or rather invisibilizes) indigenous peoples, and how it adopts mechanisms for their governance (Walter et al., 2019). Historically, colonial mechanisms of data collection and creation about indigenous people have served the interests of settler-colonial governance and disabled indigenous populations from claiming control over how they are represented in government statistics and policy. This often leads to invisibilizing and neglecting indigenous populations or acknowledging their presence in data largely as 'deficient' or problematic populations.

Indigenous data sovereignty refers to the epistemological practices relating to indigenous data, which affirm the rights of indigenous peoples to determine the means of collection, access, analysis, interpretation, management, dissemination, and re-use of data pertaining to the indigenous peoples from whom it has been derived, or to whom it relates (Kukutai & Taylor, 2015). Indigenous data sovereignty specifically responds

to the unjust and inequitable practices arising from the colonial and racist histories and contexts within which indigenous people's data have been and continue to be governed, particularly data and statistical or population-level information which informs government policy. Scholars and advocates of indigenous data sovereignty also point to the potential for Big Data technologies or 'open data' ecosystems to reproduce at scale racist and colonial underpinnings of indigenous statistics produced by settler-colonial states (Walter et al., 2020).

Indigenous data sovereignty frameworks have been adopted by indigenous people around the world, particularly in CANZUS states. One of the earliest of these was the OCAP™ (Ownership, Control, Access, and Possession) principles, which were developed by First Nations communities in Canada to claim self-determination over data used in health policy. OCAP principles consider community governance and interests within data collected about First Nations people. For example, the principle of ownership affirms that the community owns information collectively (FNIGC, 2007).

Here, ownership refers to the relationship of First Nations to their cultural knowledge and collective information. This principle states that a group owns information collectively in the same way that an individual owns his or her personal information. 'Control' refers to the principle that indigenous people and their representative organizations control research and information management practices that impact them, including the ability to decide the planning of any research process. 'Access' implies that First Nations communities must have access to their information regardless of where and how it is held, and control over how such information is accessed and by whom. 'Possession' refers to the physical custody of the information and records, as an instrumental manner in which control can be exercised (FNIGC, 2007).

Similar principles and frameworks for indigenous data sovereignty have been adopted in other contexts, including the Maiam nayri Wingara Indigenous Data Sovereignty Collective in Australia,[3] the Te Mana Raraunga Indigenous Data Sovereignty Network's Charter created to advance Maori data sovereignty in Aotearoa New Zealand,[4] and the United States Indigenous Data Sovereignty Network.[5]

Principles of indigenous data sovereignty are particularly important for their consideration of indigenous communities as an agent and holder of interests in data governance mechanisms. Realizing the principles of indigenous data sovereignty requires a shift in the dominant methodologies and practices of data governance. As identified by Walter and Suina, in their study of the Albuquerque Area Southwest Tribal Epidemiology Centre (AASTEC), realizing sovereignty over data requires that

[3] See, Maiam nayri Wingara. (n.d.). *Indigenous Data Sovereignty Everywhere…All the Time*. https://www.maiamnayriwingara.org/

[4] See, Te Mana Raraunga. (n.d.). *Our Data, Our Sovereignty, Our Future*. https://www.temanararaunga.maori.nz/

[5] See, The United States Indigenous Data Sovereignty Network. (n.d.). Collaboratory for Indigenous Data Governance. https://indigenousdatalab.org/networks/

communities are involved in creating and stewarding data about themselves, as well as ensuring that data are "reliable, valid, and useful" to the community, as defined by knowledge systems that they privilege and prioritize in particular contexts (Walter & Suina, 2019). This involves creating frameworks for communities to determine questions relating to data collection, use, and access, including: Whose data to include, or who gets counted in the community?; What is the content of data collected, and how is it beneficial to the community; and how is privacy and access to such data negotiated and decided? (Snipp, 2016).

Conceptions of 'group privacy' are an important aspect of many indigenous data sovereignty principles. Several principles and practices of indigenous data sovereignty recognize privacy as a relational concept in which the interests of the group and community are recognized and privileged, as opposed to merely individualistic notions of privacy. According to Kukutai and Cormack, indigenous privacy interests are intertwined with concepts of community, sovereignty, and self-determination, and acknowledge that the privacy interests of indigenous communities are more than merely the sum of their individual members (Kukutai & Cormack, 2020). As such, they represent an ontologically distinct method of considering and representing interests in information and data. For example, while individual privacy rights over health data might ensure that particular attributes are not used to harm an individual, they do little to protect indigenous populations from being statistically misrepresented, and subsequently, from harmful health policies which affect them.

Community-centric perspectives from indigenous data sovereignty principles have also informed government policy within indigenous territories. New Zealand's Data Strategy and Roadmap, for example, addresses the principles of the Te Mana Raraunga-Māori Data Sovereignty Network Charter.[6] In particular, it attempts to draw from the principles of *manaakitanga* (mutual responsibility towards the community and the ethical use of data) and *kaitiakitanga* (the responsibility of a guardian to responsibility steward indigenous knowledge and data for the benefit of the community).

4.3 Community Governance of Traditional Knowledge

The third model of governing community data we bring attention to, is the practices and principles relating to the governance of traditional knowledge (TK) and traditional cultural expression (TCE). The World Intellectual Property Organization defines TCE as "cultural materials created by the community that reflects that community's cultural and social identity", and TK as "traditional technical know-how, or traditional ecological, scientific or medical knowledge" (WIPO, 2003).

Community-based legal and institutional frameworks for governing the creation, use, and sharing of TK and TCE have emerged in response to dominant frameworks of intellectual

[6] See, Te Mana Raraunga. (n.d.). *Our Data, Our Sovereignty, Our Future*. https://www.temanararaunga.maori.nz/

property rights, which have been criticized as being inadequate in considering the needs of communities and failing to incorporate customary legal traditions of generating, using, and sharing access to knowledge resources.

In particular, regimes of intellectual property rights have been criticized for their inability to recognize group- or community-based rights over the forms of knowledge and information sought to be protected, and the inability to allow communities, particularly indigenous communities, to collectively exercise rights over TK and TCE. Many customary laws and conceptions of TCE and TK explicitly recognize a community or a group as having an interest in that creation or information, which are generally incompatible with liberal traditions of intellectual property rights and the dichotomy of the 'individual' intellectual property and 'public' domain (Mathiesen, 2012).

While related to socio-cultural and political critiques of contemporary information and knowledge economies, debates around TK are particularly important to study in the context of their adaptation into and relationship with existing legal and institutional contexts, for example, within frameworks of contractual law and intellectual property licenses, as well as domestic and international legal agreements. Several institutional policies, practices, and legal systems have sought to enable such forms of community-based control, recognizing particular communities as rights-holders. For example, the governments in Australia and Fiji have considered incorporating "indigenous communal moral rights" for indigenous communities to claim rights to the integrity of a work and claim attribution under copyright law (Nand, 2012).

The implementation of access and benefit-sharing mechanisms for indigenous communities under the framework of the Nagoya Protocol also provides some examples of efforts towards incorporating customary law relating to communal governance within TK and TCE mechanisms. The Nagoya Protocol is an international agreement governing genetic resources, including digitized biocultural information. The Nagoya Protocol attempts to preserve the rights of communities over TK through specific requirements for entering into mutually agreed upon terms and for prior informed consent in relation to sharing and accessing biocultural information.

There are a few instances where such mandates have been attempted to be met. The Inter-Community Agreement on Benefit Sharing, entered into with indigenous communities in Peru, and the Bailique Community Protocol in Brazil, are two examples of how community rights over TK and TCE were sought to be recognized and realized within existing legal and institutional contexts of intellectual property laws (Castro & Ramos, 2016). In both, there was a concerted effort towards participatory methodologies for conveying community rights over biocultural heritage in the form of written agreements and protocols, taking into account customary laws and traditions.

These protocols and agreements attempt to contractually oblige data governance mechanisms in line with customary law and practice. For example, according to the

Potato Park Inter-Community Agreement:

> "Everyone has the right to freely access knowledge and resources and to use them according to traditional practices and their own needs. On the other hand, they have the obligation to maintain the flow of knowledge and resources among themselves and with neighboring communities, to transmit knowledge to future generations to ensure continuity, and to protect traditional knowledge and resources from third parties. This right has an exception in the case of sacred knowledge. Only specific individuals within communities can access sacred knowledge and resources, and they have a corresponding obligation to keep that knowledge and those resources secret. Other community members have the complementary responsibility to refrain from attempting to gain access to sacred knowledge and resources" (ANDES et al., 2011).

These examples are indicative of the complex negotiations required between customary laws, practices, and interests identified by communities when recognizing community rights over information and knowledge. They also indicate the existing legal and institutional frameworks within which data governance currently operates.

5. Reflecting on Commons and Community Data Governance Practices

Community archives, indigenous data sovereignty, and the governance of traditional knowledge represent three related but conceptually distinct theories and practices of governing information and knowledge, wherein community and group rights over information have been explicitly recognized and sought to be accommodated. As data governance policy in the Global South begins to contend with issues around data sovereignty and group interests in information, it is pertinent to learn from the histories of these efforts towards conceptualizing group rights and community governance of data, information, and knowledge. This is not least because many of them have emerged specifically in post-colonial contexts in response to inequities of information and knowledge economies – a challenge that sovereign authorities are increasingly cognizant of.

First, these practices indicate the ontological and definitional issues in identifying communities as rights-bearers within liberal legal traditions of data governance, which are premised on individuals as agents and on concepts of property ownership, and provide some insight into how these might be negotiated. As noted by Anderson:

> "The politics of community arise precisely because communities are not static or bounded, but instead are dynamic and fluid. Communities come together for different purposes and in different contexts; they split, coalesce, or develop over time. The point here is that there is no clear consensus about the markers to be used in identifying a community or membership of a community. The intense politics that surround the term make its very use open to contest and dispute. Communities are notoriously difficult to define, as the abstract identification is likely to bear little resemblance to

> the practical social reality at a given space and time. The key point is that the category of 'community' is anything but stable and is thus a difficult notion upon which to rest legislative remedies" (Anderson, 2004).

Each of these practices has grappled with how to demarcate and draw boundaries around the relevant group deemed to be able to exercise group rights. Some of the ways in which these practices have attempted to negotiate questions of ownership include conscious designs which place underlying information and knowledge resources in the custody of a trusted steward, rather than requiring the transfer of ownership or material property outside of the community. These practices also align with emerging forms and notions of 'data stewardship' within data governance policy.

Commons scholarship similarly places emphasis on clearly defining the boundaries of the community that may participate in access and stewardship of a resource. In the absence of such clearly defined boundaries, outsiders with no interest in the sustainability of the resource and the community may be able to use the resource to its detriment (Ostrom, 1990). The immaterial nature of data and knowledge resources may suggest that strict definitions of community may not be necessary, at least from the perspective of preventing the exhaustion of the resource. However, clearly defining the community also helps engender trust in the institution for those choosing to comply with the agreed-upon rules, by ensuring that only those people participating in the rules are accessing the resource. In contrast with open data regimes, data access within a limited community may engender greater trust, and therefore greater sharing of data, as the community feels protected from outsiders without an interest in the resource (Beckwith, 2019). Thus, the exclusion of certain categories of users may be necessary to sustain community stewardship.

> "If data policy for the Global South is to focus on empowering communities, rather than reproducing the racist, colonial, and extractive infrastructures and practices of contemporary information capitalism, we argue that such policy should be informed by the theories and practices of information and knowledge management, which have been developed by and for communities."

Prainsack's (2019) proposal for determining unjust inclusion/exclusion from the commons may provide direction for conceiving the community in terms that enable participation while acknowledging their contingent and fluid reality. She argues that information commons should have a framework in place to determine whether certain types of exclusion (or inclusion) – from access, use, benefit, or governance – may be unjust, and negatively affect the fundamental needs and interests of those affected. The community may then apply various governance mechanisms, including legal obligations for inclusion (for contribution, access, benefit, or governance) to mitigate the possible undue harms.

Second, the practical mechanisms adopted by the professionals responsible for the design of community archives, indigenous data sovereignty methodologies, and TK and TCE agreements, can help inform the design and governance of contemporary information systems, including 'Big Data' systems. Gebru and Seo Jo, for example, identify archival practices of managing socio-cultural data – including making explicit mission statements and collection policies, enabling participation through crowd-sourcing data, and documenting and supervising data collection practices within archives – as practices which ML practitioners can and should draw from (Jo & Gebru, 2020). TK and TCE governance mechanisms have adopted methods for attaining prior informed consent and mutually agreed terms that attempt to incorporate community practices and customary laws into written agreements. Principles of indigenous data sovereignty have been incorporated into metadata organization and content management systems like Mukurtu, an open-source tool that allows indigenous communities to define levels of access and privacy controls over digitized communal heritage.[7]

McGinnis and Ostrom (2008) similarly suggest that the analytical structure of governance problems faced by local institutions may share features with governance questions at a larger scale. Community archival methodologies thereby present directions for how community-centric data governance at a larger scale may be designed.

Third, and perhaps most pertinently, these practices indicate that designing for community governance mechanisms must necessarily be bottom-up, decentralized, participatory, and contextually informed. This involves explicitly enabling communities to participate in the collection and governance of information, from deliberations about who to count amongst its members, to how to create contextual information categorization standards relevant to their particular context, and how to deliberate inter se on questions of use, valuation, and access to information. As such, monolithic or 'universal' standards for governing community data, applied equally across diverse groups, can potentially exclude community participation and agency, and risk replicating the inequities of globalized information systems (Montenegro, 2019).

Successful community stewardship is premised on rules regarding access, use, etc., suited to both the nature and characteristics of the resource, as well as the culture, ideology, and customs of the community (Ostrom, 1990). This is equally applicable to data, a seemingly global, de-territorial resource, but whose curation and use is always enmeshed in specific local configurations and contexts (Gitelman, 2013). A key objective is to maintain congruence between the effort and cost associated with participation in the collective action institutions, and the benefits users derive from it. This balance must also be perceived to be fair and equitable for all members of the community (Cox, Arnold, & Tomás, 2010). For instance, commercial entities that benefit from the resource should reciprocally contribute to the community as well. This could be in terms

[7] See, Mukurtu. (n.d.). *Welcome to Mukurtu CMS.* https://mukurtu.org/

of reciprocal obligations to contribute (data) to the common pool. Similarly, entities may be required to contribute towards necessary governance, stewardship, and technical functions. Bottom-up democratic participation is critical to ensure that rules are tailored to the local conditions of the resource and the community. The design principles stress that individuals who are affected by the rules around the resource are able to participate in the processes that make and modify the rules. Rules created by officials or authorities removed from the local context may not be sufficiently tailored, or adaptable to change over time. Regimes that empower a local elite to make the collective choice decisions face the possibility that the rules disproportionately benefit this elite. Thus, a critical feature is that the collective choice mechanisms are not undermined or co-opted by locally powerful or external bureaucratic actors (Ostrom, 1990).

In particular, the relevance of 'post-custodial' archiving aligns with the importance of accountable monitors in community stewardship. Successful commons require monitoring of resource conditions as well as compliance with the community's rules. Monitors may either be from the community, be selected by the community, or otherwise accountable for their performance (Cox, Arnold, & Tomás, 2010). They may be most effective when they benefit directly from effective management of the resource. Monitors must particularly be accountable to those in the community most dependent on the condition of the resource. These mechanisms must additionally be supported by low-cost, accessible dispute resolution mechanisms and systems or proportionate sanctions that maintain trust in the community's rules.

6. Conclusion: Institutionalizing Principles and Practices of Community Data Governance

The skewed nature of the contemporary globalized data economy is an outcome not only of the particular forms that data systems have taken, but equally of the legal, organizational, and institutional forms that comprise the totality of information infrastructure responsible for governing data. This includes laws and policies, the designs of platforms and digital information systems, and the standards and rules that govern how various components of the data economy function. We have argued that the 'community' is an important unit of analysis, first, because it may better articulate harms from contemporary data practices, and second, as a touchstone for determining beneficial data practices that meaningfully 'socialize value'. However, systematic recognition of community rights over management of their data is in many cases incompatible with existing legal and technical frameworks for data governance. We must then ask, how might principles and practices of community-based data governance be adopted into data governance policy, at the level of a polity, a digital platform, or even a technical standard?

Governments around the world are already attempting to build institutional legal and policy responses towards governing certain forms and aspects of data in ways that recognize community interests and participatory governance, moving away from the

dominant model of recognizing only individuals as relevant rights-holders.

For example, the Government of India's report on Non-Personal Data explicitly recognizes "community data" as a category of information, and outlines a framework of "data trusts" through which communities are intended to be beneficiaries of data held by "trustees" (Government of India, 2021). However, the policy lacks a coherent framework for recognizing the boundaries of self-constituted or externally recognized communities. Institutional frameworks like data trusts can draw from the literature on (digital and data) commons and community knowledge management practices we have outlined in making these determinations and thinking through appropriate mechanisms for recognizing or delineating community boundaries. This might range from models that provide informed consent to affected individuals and groups to opt-in or opt-out of data governance frameworks, to more deliberative mechanisms for identifying particular external attributes that might make up a community or a group (such as geographically bound communities like neighborhoods or cities, or legally defined communities, as indicated in the experience of indigenous communities in Australia). The traditions we explored center the community as rights-bearer, while also conceiving it as dynamic, and subject to contestation. The dynamic processes of grouping that algorithms enact, and the horizontal relationships they implicate and constitute, may similarly require not static boundaries, but contextual frameworks for determining appropriate inclusions/exclusions along them (Viljoen, 2020; Prainsack, 2019).

> "Indigenous data sovereignty specifically responds to the unjust and inequitable practices arising from the colonial and racist histories and contexts within which indigenous people's data have been and continue to be governed, particularly data and statistical or population-level information which informs government policy."

Alternatives are also being experimented with at the level of platforms or standards, to create infrastructure-level alternatives to the current dominant platforms and data governance models that rely upon centralized and hierarchical decision-making. For example, certain governments and even platforms have operationalized community-based governance models for statistical, scientific, and socio-cultural data, such as the Mukurtu Content Management System, or the Government of New Zealand's Integrated Data Infrastructure (Statistics New Zealand, 2020).

The overwhelming majority of efforts in ensuring community governance of information, data, and knowledge resources focus on the micro- or meso-level, and it is an open question as to how and whether these might be able to scale to the level of infrastructure required to systematically challenge existing models of data governance. While a full discussion of the merits or demerits of thinking about data governance at scale is not apposite, some of the practices we have spoken about do provide some insight into how community and commons-based data governance models might interact with legal or institutional systems at different scales. In particular, the history of developing sustainable access and benefit-sharing mechanisms through the Nagoya Protocol, and implementing

these mechanisms through interfaces with both community-level principles and customary data or knowledge governance mechanisms, as well as domestic legislation like contracts, patent or copyright law, gives some indication of how community data governance practices can interface with existing legal, institutional, and organizational systems operating at different scales.

In this paper, we have argued that there is value in recognizing group and community interests in certain kinds of data, an ignorance of which has contributed to the inequitable nature of the global information economy. Many of the principles and methods for the community and commons-based governance of information and knowledge resources, especially those in the Global South, have arisen distinctly in response to issues of coloniality and inequities in dominant forms of information and knowledge management in the archive, or within intellectual property laws. These practices of community data governance can, and should, inform broader ongoing policy efforts, both in recognizing the scope of rights and interests that different communities may have in managing and governing data in ways that establish agency and participation, as well as in navigating and negotiating concerns about how these rights and interests might be exercised by or within communities.

References

Agostinho, D., D'Ignazio, C., Ring, A., Thylstrup, N. B., & Veel, K. (2019). Uncertain Archives: Approaching the Unknowns, Errors, and Vulnerabilities of Big Data through Cultural Theories of the Archive. *Surveillance & Society*, *17*(3/4), 422–441. https://doi.org/10.24908/ss.v17i3/4.12330

Anderson, J. (2004). The Politics of Indigenous Knowledge: Australia's Proposed Communal Moral Rights Bill. *University of New South Wales Law Journal*. *27*(3). 585-604. http://www6.austlii.edu.au/cgi-bin/viewdoc/au/journals/UNSWLawJl/2004/34.html

ANDES, Potato Park Communities, & IIED. (2011, October). *Community Biocultural Protocols.* https://www.iied.org/sites/default/files/pdfs/migrate/G03168.pdf

Baarbé, J., Blom, M., & De Beer, J. (2017). A Data Commons for Food Security. *SSRN Electronic Journal*. https://ssrn.com/abstract=3008736

Beckwith, R., Sherry, J., & Prendergast, D. (2019). Data Flow in the Smart City: Open Data Versus the Commons. In *The Hackable City* (pp. 205–221). Springer Singapore. https://doi.org/10.1007/978-981-13-2694-3_11

Bollier, D., & Hefrich, S. (2019). *Free, Fair and Alive: The Insurgent Power of the Commons.* New Society Publishers.

Castro, C.P., & Ramos, R.P. (2016) *The Brazilian experience with community protocols and the building of prior informed consent to protect traditional knowledge*. Policy Brief OELA. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2835718

Cohen, J. (2019). *Between Truth and Power: The Legal Constructions of Information Capitalism*. Oxford University Press.

Cox, M., Arnold, G., & Tomás, S. V. (2010). A review of design principles for community-based natural resource management. In *Ecology and Society, 15*(4). Resilience Alliance. https://doi.org/10.5751/ES-03704-150438

De Angelis, M. (2017). *Omnia Sunt Communia: On the Commons and the Transformation to Postcapitalism.* Zed Books.

Delacroix, S., & Lawrence, N. D. (2019). Bottom-up data Trusts: Disturbing the 'one size fits all' approach to data governance. *International Data Privacy Law*, *9*(4), 236–252. https://doi.org/10.1093/idpl/ipz014

Evans, B. J. (2016). Barbarians at the gate: Consumer-driven health data commons and the transformation of citizen science. In *American Journal of Law and Medicine*, *42*(4), 651–685. https://doi.org/10.1177/0098858817700245

First Nations Information Governance Centre (FNIGC). (2007). *OCAP: ownership, control, access and possession, sanctioned by the First Nations Information Governance Committee*. Assembly of

First Nations, National Aboriginal Health Organization, Ottawa.

Frischmann, B., Madison, M., & Strandburg, K. (Eds.). (2014). *Governing Knowledge Commons*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199972036.001.0001.

Gitelman, L. (Ed.). (2013). *Raw Data is an Oxymoron.* MIT Press.

Government of India. (2021). *Report by the Committee of Experts on Non-Personal Data Governance Framework*.

Grondona, V., Chowdhary, A. M., & Uribe, D. (2020). *National measures on taxing the digital economy* (Research Report No. 111). Research Paper. https://www.econstor.eu/handle/10419/232232

Grossman, R. L., Heath, A., Murphy, M., Patterson, M., & Wells, W. (2016). A case for data commons: Toward data science as a service. *Computing in Science and Engineering*, *18*(5), 10–20. https://doi.org/10.1109/MCSE.2016.92

Halim, M. (2018). The Westbury Community Archive: Claiming the Past, Defining the Present towards a Better Future. *Education as Change*, *22*(2). https://doi.org/10.25159/1947-9417/3709

Harris, V. (2002). The archival sliver: Power, memory, and archives in South Africa. *Archival Science*, *2*(1), 63–86. https://doi.org/10.1007/BF02435631

Hess, C., & Ostrom, E. (Eds.). 2007. *Understanding Knowledge as Commons: From Theory to Practice.* MIT Press.

Ho, C., & Chuangt, T. (2019) Governance of communal data sharing. In A. Daly, K. Devitt, & M. Mann, (Eds.) *Good Data*. Amsterdam: Institute for Network Cultures. pp. 202-2013. https://doi.org/10.1007/978-0-387-72596-3_12

IIED, A. A. (n.d.). *Community biocultural protocols: Building mechanisms for access and benefit sharing among the communities of the Potato Park based on Quechua customary norms (summary report)*. Publications Library. Retrieved 27 September 2021, from https://pubs.iied.org/g03168

Jo, E. S., & Gebru, T. (2020). Lessons from Archives: Strategies for Collecting Sociocultural Data in Machine Learning. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 306–316. https://doi.org/10.1145/3351095.3372829

Kovaks, A., & Ranganathan, R. (n.d.). *Data sovereignty, of whom? Limits and suitability of sovereignty frameworks for data in India*. Internet Democracy Project. Retrieved 27 September 2021, from https://internetdemocracy.in/reports/data-sovereignty-of-whom

Kukutai, T., & Cormack, D. (2020). "Pushing the space": Data sovereignty and self-determination in Aotearoa NZ. In M. Walter, T. Kukutai, S. R. Carroll, & D. Rodriguez-Lonebear (Eds). *Indigenous Data Sovereignty and Policy*. Routledge.

Kukutai, T., & Taylor, J. (2016). Data sovereignty for indigenous peoples: Current practice and future needs. In T. Kukutai, T., & Taylor, J. (Eds.) *Indigenous Data Sovereignty* (pp. 1–22). ANU Press. https://www.jstor.org/stable/j.ctt1q1crgf.8

Lehtiniemi, T., & Kortesniemi, Y. (2017). Can the obstacles to privacy self-management be overcome? Exploring the consent intermediary approach. *Big Data and Society*, *4*(2). https://doi.

org/10.1177/2053951717721935

Maiam Nayri Wingara. (n.d.). Maiam Nayri Wingara. Retrieved 27 September 2021, from https://www.maiamnayriwingara.org

Mathiesen, K. (2012). A Defense of Native Americans' Rights over Their Traditional Cultural Expressions. *The American Archivist*, *75*(2), 456–481. https://doi.org/10.17723/aarc.75.2.0073888331414314

McGinnis, M., & Ostrom, E. (2008). Will lessons from small-scale social dilemmas scale up? In *New Issues and Paradigms in Research on Social Dilemmas* (pp. 189–211). Springer. https://doi.org/10.1007/978-0-387-72596-3_12

Montenegro, M. (2019). Subverting the universality of metadata standards: The TK labels as a tool to promote Indigenous data sovereignty. *Journal of Documentation*, *75*(4), 731–749. https://doi.org/10.1108/JD-08-2018-0124

Mukurtu. (n.d.). Mukurtu CMS. Retrieved 27 September 2021, from https://mukurtu.org/

Nand, S. (2012). Protecting and revitalizing traditional knowledge and expressions of culture: For an equitable future in Fiji. In A. Mendes-Walsh & S. Meghana (Eds.) *Research papers from The WIPO-WTO Colloquium for Teachers of Intellectual Property Law - Geneva, 2011*. (pp. 55–73). Geneva: World Intellectual Property Organization.

Ostrom, E. (1990). *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge: Cambridge University Press. http://doi.org/10.1017/CBO9780511807763

Poole, A. H. (2020). The information work of community archives: A systematic literature review. *Journal of Documentation*, *76*(3), 657–687. https://doi.org/10.1108/JD-07-2019-0140

Prainsack, B. (2019). Logged out: Ownership, exclusion and public value in the digital data and information commons. *Big Data and Society*, *6*(1), 205395171982977. https://doi.org/10.1177/2053951719829773.

Raine, S.C., Kukutai, T., Walter, M., Figueroa-Rodrigues, O.L., Walker, J. et al. (2019). Indigenous data sovereignty. In T. Davies, S. Walker, M. Rubinstein, & F. Perini (Eds.), *State of Open Data: Histories and Horizons* (pp. 300–319). Cape Town: African Minds.

Snipp, C. M. (2016). What does data sovereignty imply: what does it look like? In T. Kukutai & J. Taylor (Eds.), *Indigenous Data Sovereignty: Toward an agenda* (pp. 39–56). ANU Press. http://www.jstor.org/stable/j.ctt1q1crgf.10

Statistics New Zealand. (2020). *Integrated Data Infrastructure*. https://www.stats.govt.nz/integrated-data/integrated-data-infrastructure.

Taylor, L. (2017). Safety in Numbers? Group Privacy and Big Data Analytics in the Developing World. In *Group Privacy*. Springer International Publishing. https://doi.org/10.1007/978-3-319-46608-8_2

Te Mana Raraunga. (n.d.). Te Mana Raraunga. Retrieved 27 September 2021, from https://www.temanararaunga.maori.nz

The World's Most Valuable Resource is no Longer Oil, but Data. (2017, May 6). *The Economist*.

https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data

United States Indigenous Data Sovereignty Network. (n.d.). United States Indigenous Data Sovereignty Network. Retrieved 27 September 2021, from https://usindigenousdata.org

Viljoen, S. (2020). Democratic Data: A Relational Theory For Data Governance. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3727562

Vipra, J. (2019, February 28). *Protect people's economic rights to data, RCEP countries urged—Focus on the Global South*. https://focusweb.org/protect-peoples-economic-rights-to-data-rcep-urged/

Walter, M., & Carroll, S. R. (2020). Indigenous Data Sovereignty, governance and the link to Indigenous policy. In M. Walter, T. Kukutai, S. R. Carroll, & D. Rodriguez-Lonebear (Eds). *Indigenous Data Sovereignty and Policy*. Routledge.

Walter, M., & Suina, M. (2019). Indigenous data, indigenous methodologies and indigenous data sovereignty. *International Journal of Social Research Methodology*, *22*(3), 233–243. https://doi.org/10.1080/13645579.2018.1531228

Walter, M., Lovett, R., Maher, B., Williamson, B., Prehn, J., Bodkin-Andrews, G., & Lee, V. (2021). Indigenous Data Sovereignty in the Era of Big Data and Open Data. *Australian Journal of Social Issues*, *56*(2), 143–156. https://doi.org/10.1002/ajs4.141

World Intellectual Property Organization. (2003). *Traditional knowledge, traditional cultural expressions and genetic resources: The international dimensions*. Intergovernmental Committee on Intellectual Property and Genetic Resources, Traditional Knowledge and Folklore. Geneva: WIPO, WIPO/GRTKF/IC/6/6. https://www.wipo.int/edocs/mdocs/tk/en/wipo_grtkf_ic_6/wipo_grtkf_ic_6_6.pdf

Wurl, J. (2005). *Ethnicity as Provenance: In Search of Values and Principles for Documenting the Immigrant Experience*. https://minds.wisconsin.edu/handle/1793/45926

Bambauer, J. (2012). Tragedy of the Data Commons. *Harvard Journal of Law and Technology*. https://doi.org/10.2139/ssrn.1789749