11

ABS: Big Data, Data Sovereignty and Digitization

A New Indigenous Research Landscape

Chidi Oguamanam

Abstract

This chapter focuses on the increasing sophistication of research practices through the applications of digitization and other aspects of information and communication technology (ICT). Multiple factors, including advances in biotechnology and the production, utilization and malleability of valuable research data through the use of digital technology tools have resulted in the transformation of data or genetic information into widely accessible virtual resources that are practically de-linked from their origins. Given the orientation of the Nagoya Protocol towards the physical transfer of genetic resources, the virtualization of Indigenous research data makes the latter part of the big and open data grab threatening the realization of ABS. However, despite the potential to de-link genetic resources (GRs) and associated traditional knowledge (aTK), including other aspects of Indigenous research data from their sources, conceivably, there are significant bases in the texts of CBD and the Nagoya Protocol for the inclusion of digitally sequenced data as part of ABS. Further, the interface of Indigenous peoples and local communities' (IPLCs) nascent interest in data sovereignty and the big and open data phenomena provide an opportunity to apply critical data analytics to mainstream data equity as an integral aspect of Indigenous-sensitive ABS in an increasingly sophisticated and technology-driven research environment.

INTRODUCTION

Article 1 of the Nagoya Protocol states, '[t]he objective of this Protocol is the fair and equitable sharing of the benefits arising from the utilization of genetic resources, including by appropriate access to genetic resources . . .'. This provision partially

https://doi.org/10.1017/9781108557122.012 Published online by Cambridge University Press

repeats Article 1 of the CBD. The essential focus of the Nagoya Protocol (NP) is GR, defined in its parent convention as 'genetic material of actual and potential value' (CBD, 1993, Art 2). In order to trigger claims for equitable benefit sharing, there must be in effect 'utilization of genetic resources.' The NP, as opposed to the CBD, defines utilization of GR as the 'conduct of research and development on the genetic and/or biochemical composition of genetic resources, including through the application of biotechnology as defined in Article 2 of the Convention' (NP, Art 2(c)). The CBD defines biotechnology as 'any technological application that uses biological applications, living organisms and derivatives thereof, to make or modify products or processes for specific use' (NP, 2010, Art 2(d); CBD, 1993, Art 2). The protocol defines derivatives as 'naturally occurring biochemical compound[s] *resulting from the genetic expression or metabolism* of biological or genetic resources, even if [they do] not contain functional units of heredity'[1] (NP, 2010, Art 2(e)). Also, as at June 3, 2016 the WIPO-IGC consolidated document to IP and GRs adopts that same definition of derivatives.[2]

In the structure of the both the CBD and the NP, the focus is essentially on GR. No mention is made of TK until Article 8(j), in the case of the CBD, and Articles 7, 10, 11, 12, 13 etc., in the case of NP, which make reference to 'traditional knowledge associated with genetic resources.' There is no definition of that concept in either the CBD or the NP (see Phillips, Smyth & de Beer, Chapter 10). WIPO-IGC's attempt to define TK and TK associated with GR remains inchoate. From the above perspective on utilization, it is logical to assume that any research[3] that involves TK associated with GRs or vice versa, as the case may be, amounts to the utilization of the TK and the GRs and, consequently, triggers equitable ABS claims. As tripartite concepts, research, TK and GRs inherently derive their relevance not only as blurry forms of datasets or information in and of themselves, but also in the production and utilization of datasets and information.

Both the CBD and the NP are silent on 'derivatives' as they apply to TK. But a strict textual appraisal of the language of the NP would suggest that it does not discount the notion of derivatives in relation to TK. First, it appears that the moment TK is associated with GRs, the latter becomes susceptible to the provision on derivatives as outlined above. Second, even if the first proposition is shaky, which is not conceded, the NP is consistent in its text to the effect that, TK associated with GR held by IPLC must be accessed 'with their prior informed consent' (PIC) or with their approval and involvement pursuant to mutually agreed terms (MAT) (NP, 2010, Art 7, 11, 12, 13, 16, 18, etc.). The NP leaves wide discretion for parties regarding how the involvement, approval and prior informed consent of the IPLCs could be secured and with regard to the constitution of MAT.

Though these mutually reinforcing provisions are within the ambit of progressive evolution of international law on Indigenous peoples as echoed in the NP preamble, IPLCs are free to articulate and ensure that the derivatives of TK are subjected to equitable ABS in ways that go beyond the narrow confines of the definition of

https://doi.org/10.1017/9781108557122.012 Published online by Cambridge University Press

derivatives proffered above. The CBD preamble recognizes the malleability of the applications and manifestations of TK beyond the scientific reference to derivatives as genetic expression or metabolism of naturally occurring biochemical compounds. Specifically, the CBD recognizes 'the unique circumstances where TK associated with GRs is held in countries, which may be oral, documented or in other forms, reflecting a rich cultural heritage for conservation and sustainable use of biological diversity' (CBD, 1993, Art 23).[4] Simply put, when GRs are associated with TK or when TK is associated with GRs, the manifestations of that TK may not necessarily be linked to *physical* representations. Rather, it may involve nuanced forms that go beyond the emphasis of the texts of both the NP and the CBD on physical transfer of GRs and, as the case may be, associated TK.

Despite the recognition of the character of TK in the CBD, the latter and its NP focus essentially on the corporal notion of GRs. For example, both recognize the sovereign rights of nations to exploit their own natural resources (CBD, 1993, Arts 3, 15; NP, 2010, Art 6). As well, GRs are designated in *ex-situ* and *in-situ* terms. The provisions on practical implementation of the NP make references to transboundary GRs and transboundary collaboration as well as to checkpoints for monitoring of compliance (NP, 2010. Arts 10, 11, 14, 17). Logically, these are references to recourses in their physical forms. For the most part, some of these provisions reflect a bias for the physical character and expectations for the physical movement or transfer of GRs under the ABS system.

However, technological reality has forced the CBD to grapple with digital sequence technologies as they apply to genetic resources. This is as a consequence of the fact that the utilizations of GRs and/or TK associated with GRs are happening in contexts devoid of their physical transfer or movement. In the information age, research is essentially a data-driven initiative, animated by open-ended possibilities for the generation, manipulation, diffusion, anonymization and various innumerable forms of data aggregations (see Phillips, Smyth & de Beer, Chapter 10). On a positive side, digital technology has lowered the cost of data; enhanced the accessibility and exchange of vital information in ways that facilitate research and promotes its objectives of advancing knowledge, proffering solutions to problems, and ultimately improving human capacity and quality of life. The prevalence of technology in the generation and management of data therefore creates both opportunities and dilemmas for stakeholders in the context of the interface between GRs and TK, with significant implications for ABS.

In a recently commissioned study by the CBD on digital sequence information on GR, experts point out that the interaction of these phenomena represents a profound area that shapes contemporary research. It constitutes a significant challenge to the implementation of the NP ABS scheme especially with regard to the identification of contributors or users of GR and associated traditional knowledge and provenance of sequences. This chapter explores that important dynamic – the understanding and mitigation of which is critical for an Indigenous-sensitive ABS in Canada and around the world.

https://doi.org/10.1017/9781108557122.012 Published online by Cambridge University Press

THE BIG DATA PHENOMENON

Over the last two decades, there has been an explosion of interest in the concept of 'big data,' initially by technology companies and data-based giants such as Google, Facebook, Twitter, Instagram, eBay, Amazon and Wikipedia, to name just a few. Big data designates the phenomenon of massive and complex data sets at a scale at which it is not possible for conventional data processing applications to handle. Because of their richness in information, these massive datasets have been turned into goldmines for the application of predictive analytics, user behaviour analytics and other sophisticated data analytical methods in order to harvest or extract insights and optimize the unprecedented value in the novel data ecosystem. Big data has since been of significant interest across many areas of human endeavour, including social behaviours, environment, marketing, manufacturing, healthcare, DNA mapping or sequencing and profiling (Oguamanam, Chapter 14; Phillips, Smyth & de Beer, Chapter 10) education, and governance, to mention a few. Public and private sectors in the United States, Canada, United Kingdom, the European Union, China and India have continued to invest in big data as a part of their intense competition to leverage the information and communication technologies in virtually all sectors of human endeavour.

One of the most profound applications of big data is in the realm of research. Big data rapidly generates vital research information that is usable in a variety of disciplines. For example, it took ten years for the Human Genome Project (HGP) to decode the human genome; with big data, the same feat now takes just a single day. Big data reduces the cost of research. Again, in regard to decoding the human genome, using big data, the cost has been reduced by over 100 times. Big data enhances the generation and storage of information across distance and time, including those relating to genetics, genomics, biomes, biological properties, environment, climate and geology, consumer behaviours, historic patterns or phenomenon, etc. to rapidly advance social, commercial, and health experiments and interventions. It facilitates and entrenches a culture of open repository of vital research information through multiple information pulling applications (wireless devices, networked sensors, aerial sensors, cloud computing, RFIDs, etc.) to be easily accessed by researchers at minimal or no costs.

BIG DATA, OPEN DATA AND OPENNESS

As an adjunct of the new information technology era, big data is an important catalytic and incentivizing factor to openness, open innovation and open source and open data. Like the concept of openness, big data is not necessarily antithetical to proprietary use of data or information. In fact, private sector corporations largely drive the big data phenomenon for their firms' needs as a competition and survival strategy. Both big data and open data or the open source phenomena generally are constructive and modified forms of proprietary use of data in self-interested ways that

strategically encourages targeted forms of sharing via licencing or related schemes to optimize value (de Beer, 2016; Phillips, Smyth & de Beer, 2017). Notwithstanding the corporate proprietary interests that have been pivotal in the evolution of big data, the latter have significant effects in promoting access and democratising the use and advancement of the impact of large scale information on the society. As such, big data has a nuanced relationship with open data and open source. In Canada, the University of Waterloo runs the Canadian Open Data Experience (CODE). The program demonstrates the relationship between open and big data, including the role of open data (specifically data visualization) in the advancement of big data.

'DATAMANIA' AND RESEARCH INVOLVING INDIGENOUS PEOPLE

As with other sectors, research involving Indigenous peoples is not immune from the effects of digital technology, big data and open data and their applications in the processing of sensitive data like genetic/genomic, ecological information, or even patterns, demography or the mapping and various nuances of traditional cultural practices. We called attention to this tendency in an earlier work where we observed as follows:

> Multidisciplinary researchers ranging from cartographers, ethnographers, anthropologists, economists, social scientists, critical data studies experts to lawyers doing traditional knowledge-related research by and with Indigenous communities 'have witnessed the emergence of numerous issues regarding the collection, dissemination and management of data based on Traditional Knowledge.'[5] Not only do such issues implicate the problematic relationship between intellectual property and traditional knowledge, they also touch on the subject of access and equitable sharing of benefits arising from such research.
>
> (Oguamanam & Jain, 2017, 95)

Nowadays, research involving Indigenous peoples invariably results in a significant digital footprint or digital output, including through online data resources of various forms such as text, images, audio, video, data versioning mapping, etc. Researchers, including those involved with Indigenous peoples commonly establish dedicated webpages for their projects. These web platforms are proven sources of significant, publicly accessible data which can be mined and interpreted as part of the global universe of big data without the knowledge nor the guarantee of compliance with the terms of engagement between Indigenous peoples and researchers. Despite any stated conditions and caveats, which are mostly unenforceable, any such independent access of often vital Indigenous research-related data is open to further de-contextualization and (mis)interpretation without recourse to Indigenous peoples.

The continued pre-eminence of the big data and open data movements results in availability of an unprecedented scale of Indigenous research-related data that are conceivably inseparable from TK, for the most part, at virtually all levels of the use of GRs in traditional knowledge innovation and practices (TKIP) including but not limited to health, traditional medicine, agriculture, cultural expressions, sacred and secret rituals, food, genomics, culture, social behaviour, demographics, ethnography, climate management, hunting, special environments such as the polar region (Scassa & Taylor, 2017), etc. Not only are these kinds of information easily de-linked from their sources and origins in ILC. In their transformations, they may or may not result from direct physical dealings with GRs. But because they are parts of research and consequently constitute utilizations of GRs and, as may be applicable, aTK, they are, arguably, subjects of ABS obligations in accordance with the specific provisions of the NP and the CBD examined in the earlier section of the chapter and elaborated further below in the discussion on derivatives.

In Canada, the Geomatics and Cartographic Research Centre (GCRC) at Ottawa's Carleton University presents a variant form of approach and a new form of experience over the increasing web presence and digital footprints of data arising in the context of Indigenous research. The GCRC works in partnership with Northern Canadian Indigenous communities deploying geographic processing and management skills as predicative and other tools of analysis 'for a range of socio-economic issues of interests with a focus on specific local and international contexts.'[6] The project is partly driven by an online interactive atlas on geographic, geomatics, cartographic, environmental and TK practices [of Indigenous peoples] developed by the Centre. Other related Indigenous research endeavours that directly and indirectly project TK and associated data into cyberspace in furtherance of the intersection between open and big data are prevalent in Canada and elsewhere (e.g. Mapping in Indigenous Communities Project). Scassa and Taylor (2017) have recently broached the issue of ethical challenges for the inclusion of TK as part of the Arctic data infrastructure.

GCRC researchers recognize that third parties who seek access to TK are often driven by commercial and intellectual property needs. Given the reductionist nature of those prisms in relation to the 'communal and other unique features of traditional knowledge' (Oguamanam & Jain, 2017, 95) they proposed an *open* licencing scheme for TK with the objective of assisting 'traditional knowledge holders communicate their expectations for appropriate use of their knowledge to all end users – a development that potentially contributes to the letter and spirit of ABS and to other non-economic aspects of traditional knowledge' (Oguamanam & Jain, 2017, 95). While a licencing scheme for TK, as a contractual matter, has potential to accommodate Indigenous peoples' ABS sensitivities and more, its viability may be contingent on several contextual variables, including the nature of the TK or GRs, where applicable, and the dynamics of a given Indigenous community as well as the envisaged use for the GRs and TK. One can confidently suggest that the GCRC TK licencing scheme is truly an 'open' proposition in both a literal and figurative sense.

The GCRC's research model reflects an attempt to temper the big and open data imperative with a sensitivity that is historically demanded by Indigenous peoples over dealings with their TK and associated GRs. But that is only one research initiative. A more systematic engagement is required to square up the open and big data phenomena with Indigenous peoples' expectations over the sourcing and use of their data in the new research environment in ways that address equity deficits in the use of TK (de Beer, 2016). That very imperative, which is captured by Indigenous peoples' interest in data sovereignty, explored below, is critical to the development of an Indigenous-sensitive domestic ABS policy in Canada. In addition to University of Waterloo's CODE mentioned earlier, in Canada, the Open North initiative is another open data program and part of the global big data movement that promotes the use of civic technology tools at both domestic and global levels to foster the public goods and democratic benefits of big data. With its focus on the First Nations of Canada's North, the Open North initiative inevitably engages Indigenous peoples' inclination toward data sovereignty with its commitment to open data (Lauriault, 2017; Oguamanam & Jain, 2017).

CASE FOR DIGITAL DNA

Before turning to the meaning and rationale for Indigenous peoples' quest for data sovereignty, an important but obvious point deserves a brief mention. The global big data and open data phenomena's role in the virtualization, malleability and democratization of data access for complex objectives is a factor or spinoff of information communication technology. But beyond ICTs, advances in biotechnologies constitute interrelated but additional site for the generation of critical research data in forms that de-link them from naturally occurring GRs, blurring, distancing or complicating authentic claims to their origin or source, not to mention their association with TK of Indigenous and local communities (Phillips, Smyth & de Beer, Chapter 10). For example, through multidisciplinary insights ranging from engineering, molecular and synthetic biology, chemistry to genomics, genetic epidemiology, biotechnological insights are augmented to generate various kinds of information and data relating to GRs. In some cases, the undergirding research is inspired by Indigenous knowledge of the uses of plants, animal and other genetic materials for medicinal, therapeutic, pharmacological, food and agricultural practices and innovations. In other cases, such as the Human Genome Diversity Project (Coombe & Amani, 2005) or the *map-my-gut* initiative (Spector, 2017), Indigenous peoples themselves are the sources of vital genetic material. In yet others, their traditional dietary practices are foundations of vital information and insights for understanding health-improving life practices and positive but complex human-nature interactions, for example, pursuant to the gut biomes research.[7]

Biotechnology facilitates the generation of genetic data or information including, for example, through DNA sequencing with results digitally stored in the form of digital DNA. The latter is critical for the conduct of synthetic biology research. As a

https://doi.org/10.1017/9781108557122.012 Published online by Cambridge University Press

cognate of genetic engineering, synthetic biologists are able to construct new DNA components that are not naturally occurring. They are also able to re-design existing biological forms or their properties with modular DNA parts, re-arranging and combining them in new ways that result in new and complex biological systems in predictable and well characterised manners. All of these feats of ingenuity in the interface of biotechnology and digital technology translate into outcomes that solve practical problems in innumerable range of fields. Again, in an earlier project we surmised as follows:

> Digital DNA makes it easier to conduct research. Rather than sourcing genetic sequences in nature, researchers can use online databases to download DNA sequences for free with a click of a button. These sequences can be customized and then ordered from commercial laboratories to conduct research, allowing entire genomes or genes to be constructed from scratch. As DNA synthesis and sequencing technologies become cheaper, it may be faster to synthesize certain DNA sequences than to find them in nature . . . Despite the obvious advantages of using digital DNA for research, it raises concerns for biopiracy. Users can benefit from genetic resources or local knowledge available on the web without necessarily being obliged to share the benefits derived from using the online data.
>
> (Oguamanam & Jain, 2017, 106, 107)

The point here is that a combination of ICTs' preeminent role in fuelling the big and open data phenomena, and continuing advances in biotechnology in various directions de-emphasizes the physical transfer of GRs as fundamental triggers of ABS. In order to leverage the possibilities under emerging ABS regimes whether within or without the Nagoya framework, there is need for strategic vigilance, expertise, capacity building and awareness-raising in the complex forms in which GRs and associated TK are being generated, used or transferred in biotechnological applications (Oguamanam & Hunka, Chapter 3). But the key question is whether the apparent de-linking or blurring of the sources or origins of GRs under these complex and hi-tech forms for their use and transformations compromises their status as derivatives or not. Even though the foregoing analysis and from the specific reference to derivatives in the text of the NP suggests an affirmative response. But such an inclination is not absolute. It must be mindful that the CBD Ad Hoc Technical Working Group on Digital Sequence Information (DSI) is divided on the question of whether DSI is included in the definition of GR or not (Oguamanam, Chapter 14). That dissonance among experts is related to the existing controversy surrounding the scope of derivatives among stakeholders in ABS.

DERIVATIVES AND ABS

To further buttress the case for applicability of ABS in the context of migrations of Indigenous research-driven information or data in GRs and TK to the realms of big and open data, we call attention to a 2017 World Intellectual Property Organization

https://doi.org/10.1017/9781108557122.012 Published online by Cambridge University Press

(WIPO) study, *Key Questions on Patent Disclosure Requirements for Genetic Resources and Traditional Knowledge* (WIPO, 2017). The study was conducted at the instance of WIPO's specialist Intergovernmental Committee on Intellectual Property and Genetic Resources, Traditional Knowledge and Traditional Cultural Expressions (IGC), tasked to develop texted-based instruments for the protection of those subject matters for WIPO member states. An issue at the IGC deliberations, and certainly in all cognate fora is whether and to what degree could GRs or materials and TK be an integral part of an invention to warrant the disclosure of their sources or origins in patent applications. As a practical matter, this kind of disclosure would foster accountability and transparency over the uses of GRs, enhance ABS, while ensuring that the patent system ceases to serve as a conduit for unlawful appropriation of GRs and TK.

The above-mentioned WIPO study focuses on the degree of linkage or relationship between GRs and, where applicable, TK with a claimed patent invention that could trigger disclosure of source or origin and consequently an Indigenous and local community claim for ABS. This is important because not every casual or inconsequential nexus or association between a claimed invention and GRs or TK of ILCs could be subject of ABS and disclosure claims. The identified links provide further insight for understanding and supporting the position that a generous interpretation of derivatives or process of derivation is necessary for an Indigenous-sensitive ABS policy. As well, it is already noted that the text of the NP accommodates the dynamic character of GRs and, conceivably, TK to account for the malleability and migrations of research data based on them. A 2010 Norwegian Fridtjof Nansen Institute report endorsed a 'broad' definition of GRs that reflects the 'dynamic' understanding of the concept (Fridtjot Nansen Institute, 2010; Oguamanam & Jain, 2017).[8] As well, members of the Indigenous caucuses to negotiations in cognate fora to the IGC featuring patent disclosure, protection of TK and ABS as part of the broader law and policy discourse on intellectual property and development strongly expressed a similar inclination on the issue of derivatives of GRs and TK (Bagely & Rai, 2014).

The WIPO study identified three contexts or degrees in which an invention may be linked to GRs and TK in order to warrant patent disclosure of source or origin. First, when the invention directly claims to have *utilized* GRs or TK. This is straightforward on its face. It calls attention to the definition of utilization of GRs under the NP reproduced earlier in this chapter. Second, where the invention is *derived* from GRs or TK. The third relates to where the invention is *based* directly on GRs or TK. While the third context recognizes situations of obvious and perhaps unequivocal nexus, on evidentiary bases, between an invention and GRs and TK, the second one makes reference to when an invention results as a derivative of GR and TK. It leaves open considerations for the quantity, content and quality of the role or effect of GRs and TK of Indigenous and local communities on the innovation or invention for which a patent claim is made. It is conceivable that in these three sites, the nexus between claimed invention and GRs or TK may overlap. They

are not exclusive. An invention may be based on GRs and may as well have been derived from it and in which case it inherently utilizes GRs. All of these are matter of degree and analytical disaggregation. Detailing of legal specificities and their consequences, it is argued, ought to be a domestic matter that requires Indigenous peoples' participation at policy making and implementation levels.

In making such an evaluation, it is suggested that the utilization or presence of GRs and TK be appraised on account not only of their physical transfer or physical contact with the user but also on the basis of the latter's sourcing of valuable data or information on the GRs through big and open data and other publicly accessible, including digital and web-based, platforms that have the tendency to de-link GRs and TK from their origin. Nonetheless, it is recognized that depending on the level or extent to which the accessible information constitutes the resulting invention, establishing the novelty of the invention may be problematic. This interpretive pathway is consistent with a combined reading of the NP's definitions of utilization of GR, biotechnology and derivative reproduced earlier in this chapter. Objection may be taken to the narrowness of the definition of derivative, which bears repeating: 'naturally occurring biochemical compound resulting from genetic expression or metabolism of biochemical or genetic resources even it does not contain functional units of heredity' (NP, 2010, Art 2(e)). Despite the fact that in many IPLCs' worldviews TK and GRs are part of symbiotic holism of the natural order, this definition limits derivatives to 'naturally occurring biochemical compounds' which is the dominant context for application and uses of some but not all forms of TK in association with GRs. It may be argued that the reference to naturally occurring compounds is specific to a physical object. But true as that may be, its 'utilization' broadly construed could include any and other forms in which these naturally occurring are utilized or applied, which will include how they are expressed as abstract datasets or sequences in biotechnology research.

Besides, other aspects of TK and practices based, for example, on rituals, ceremonies, protocols, etc. which are not captured in the NP are often documented in open and publicly accessible big data platforms. Their appropriations for insights by researchers and users of GRs conceivably fall well within the notion of derivatives especially when those insights are obtained through a universe of big data that elaborate the transformations or practical metamorphoses of GRs into naturally occurring biochemical compounds. That it does not matter whether or not the naturally occurring biochemical compounds have functional units of heredity is instructive. At the very least, it opens the way for sanctioning Indigenous peoples' claims for ABS over their GRs and associated TK in the realm of synthetic biology.

The NP specifies that 'utilization' of GR happens essentially in the conduct of R&D on GRs through the applications of biotechnology. There is no question that open and big data phenomena today constitute part of the most resourceful infrastructures for the conduct of research. And to the extent that GRs and aTK are increasingly becoming part of the global big data infrastructure, despite their

tendency to de-link GRs and aTK from their sources and origins in ILCs, it does not disentitle Indigenous stakeholders from making legitimate claims for ABS. Finally, biotechnology is characterized in the NP as technological application in the use of 'biological systems, living organisms or derivatives thereof to make or modify products or processes for specific use' (2011, Article 2(d)). Again, it has been noted that the interface of digital technology and continuing advances in biotechnology have boosted the uptake of big data, especially in life sciences R&D. Through interdisciplinary concerts in genetic engineering, molecular and synthetic biology, bioinformatics, genomics, genetic epidemiology, etc. research data relating to Indigenous peoples and in some cases their knowledge systems, their GRs, genetic profiles and their 'derivatives,' can readily be generated, modified or adapted to accomplish R&D objectives. We have made reference to such outcome in our discourse of digital DNA above and elsewhere (Oguamanam & Jain, 2017).

Notwithstanding the tendency of these technologies to de-link or conflate ensuing critical research data or information from their origins, the proposition for building new biological systems (with or without functional units of heredity) from scratch is dubious. This is even more so when such a system is claimed to have an absolute disconnect from natural sources or absent some form of inspiration even in regard to pattern, characteristic or predictability from those said to be 'naturally occurring,' a term that is problematic on its face. In sum, contrary to the apprehension that the migrations, malleability, dilutions and de-linking of crucial research data and information from IPLC over GRs and TK into the universe of open and big data, there is still a solid and legally sustainable case for ABS in those contexts. Yet, the overall inequitable effect of big and open data on IPLCs should neither be undermined in their entirety, nor should the flaw in the NP to directly pre-empt or accommodate the ABS implications of digital technology or, more technically, 'digital sequence information' for TK be downplayed. A more rigorous philosophical and yet pragmatic response to big data is required. It is a response that attempts to capture Indigenous peoples and, as may be applicable, other local communities' complex interests in ABS. As indicated earlier, such interests transcend the mere commercial and market value of GRs and TK. They are captured under the nascent or emergent concept of data sovereignty as it applies to Indigenous peoples in research contexts, which is the focus of the next section.

DATA SOVEREIGNTY

Despite all the benefits claimed for big and open data, there are significant degrees of skepticism around them. Such reservations are the preoccupation of a comparatively nascent field known as critical data studies; they are not of direct interest to this chapter. However, from the above analysis, it is clear the big data phenomenon is a significant factor in the de-linking, de-contextualization and virtualization of data. In relation to data arising from research dealing with GRs and TK, it has been argued

that a combination of big data and advances in biotechnology as symbolized, for example, by digital DNA and synthetic biology applications is capable of complicating claims for ABS on a practical level. As the effects of big data on all stakeholders, especially the most vulnerable, such as Indigenous peoples, attract interest of policy makers and critical data analysts, some have called for mediating such effects through some form of social contract-oriented intervention to protect vulnerable interest or values such as individual or civil liberties and privacy rights, etc. (Al-Rodhan, 2014). Indigenous peoples have articulated such interventions in the form of data sovereignty. While the social contract model and sovereignty approach are not necessarily synonymous, both could advance the course of justice, fairness and equity. Data sovereignty is arguably not a strict counterpoise to big data, but it could serve to moderate its negative effects and help explore and contextualize Indigenous vulnerabilities over big and open data phenomena and, in the presence case, with regard to safeguarding the progress made around ABS.

In its general construct, data sovereignty designates the right of States in relation to others States to govern the collection and ownership, including access and use of data that is domiciled within their jurisdiction. As well, data sovereignty denotes the sanctity or integrity of data. It is therefore an incidence of the sovereign right of States as it extends and applies to data governance. The capacity of ICTs to digitize information and strip it of any jurisdictional affiliation, for example, through cloud computing, does not fully deprive data of standing especially with regard to the locus of its storage or generation. It is only logical that the States that have the highest aggregation of contact with specific data assert sovereignty, especially pursuant, analogously, to conflict of laws principles, but subject, of course, to any contractual obligations and principles of collaboration that are critical for law and order in cyberspace.

Another aspect of data sovereignty that is relevant relates to the application of the elements of its logic onto the milieu of Indigenous peoples, with specific regard to the research context. On a more serious rendition, assertion of data sovereignty by Indigenous peoples is an aspect of their fundamental right to self-determination and their claim to shared sovereignty within collaborative federalism. In 2017, a group called the International Indigenous Data Sovereignty issued the Indigenous Charter Statement. The Group comprises three networks of Indigenous peoples organized at national levels, namely the Te Mana Raraunga - Maori Data Sovereignty Network, the United States Indigenous Data Sovereignty Network (USIDSN), and the Maiamnayri Wingara Aboriginal and Torres Strait Islander Data Sovereignty Group in Australia. Increasingly, these initiatives are promoting international Indigenous consciousness on data sovereignty reaching out to Hawaii, Lapland and other Indigenous peoples with commitment to fashioning policies on how best to collaborate in the control, sharing and application of information or data relating to research involving Indigenous peoples. As its overarching objective, Indigenous Data Sovereignty (ID-Sov for short) aspires toward 'a more robust and coherent

https://doi.org/10.1017/9781108557122.012 Published online by Cambridge University Press

international collaboration to achieve impactful outcomes at the intersection of Indigenous data sovereignty, Indigenous data governance and research' (International Indigenous Data Sovereignty IG Charter Statement, 2017). Indigenous data sovereignty concerns the rights of Indigenous peoples or nations to govern the collection and ownership, including access and use of Indigenous-related data in research and other contexts.

The movement, which is now known as the International Indigenous Data Sovereignty Interest Group (IDSIG) is committed to fostering data-driven research, promoting the use of data, building capacity in data generation within and outside the academic research contexts prioritizing benefitting Indigenous communities. The IDSIG captures the relevance of data sovereignty for Indigenous peoples in the context of big data and the role of data in changing research dynamic in the 21st century in the following statements culled from its Charter Statement. It merits significant attention:

> Like other nation states, Indigenous nations need data about their citizens and communities to make informed decisions. However, the information that Indigenous nations have access to is often unreliable, inaccurate, and irrelevant. Federal, state, and local governments have primarily collected these data for their own use. Indigenous nations' reliance on external data that do not reflect the community's needs, priorities, and self-conceptions is a threat to self-determination. The demand for Indigenous data is increasing as Indigenous nations and communities engage in economic, social, and cultural development on an unprecedented level. Given the billions of dollars in research funding spent each year and the increasing momentum of the international big data and open data movements, Indigenous nations and communities are uniquely positioned to claim a seat at the table to ensure Indigenous peoples are directly involved in efforts to promote data equity in Indigenous communities.

The Canadian situation is no different from the international context that the IDSIG articulated above. As far back as the 1999 *Royal Commission on Aboriginal Peoples*, Indigenous peoples decried the historically inequitable relationship between them and researchers. Part of the struggle to release their status as stakeholders in research involving them was the establishment of the First Nations Information Governance Centre (FNIGC) which exercises custody and control over First Nations Regional Health Surveys (RHS) data. To its credit, the FNIGC developed an RHS code of ethics, which outlines guiding principles, ethical practices and protocols for the use of data generated pursuant to the RHS. These principles link Indigenous peoples' interests in research data to their self-determination rights, which are enhanced when they benefit from the result of such research and are empowered to take control of their health and are able to receive research funds and be proactively involved in participating in research about their own peoples.

In addition to the undergirding *raison d'être* for the RHS, in 1998, the FNIGC initiative developed what, in retrospect, amounts to the first major Indigenous data

sovereignty initiative in Canada titled the OCAP principles. The acronym stands for Ownership, Control, Access and Possession. As a typology of data sovereignty, OCAP expresses the core framework for ensuring that Indigenous peoples have control over the various dealings in research data relating to them from their collection, uses or applications, dissemination, sharing (perhaps including, arguably, interpretation). All of these are supposed to happen under an ethical consciousness to protect and practically translate or give effect to those key words. We have noted, in an earlier work, that both the RHS and the OCAP principles are limited in two respects. First is in terms of the research context and the second is with regard to the category of Indigenous peoples to which they apply. Historically, they arose in the context of health-related Indigenous research data and they apply to First Nations and Inuit as opposed to other Aboriginal categories.

Those limitations are no longer valid. Both the RHS and OCAP support development of research partnerships with all researchers without limitation to any discipline or field of research. There is now in Canada an elaborate research ethics framework focused on doing various forms of research relating to Indigenous peoples as a whole (Tri-Council Policy Statement 2, 2014, Chapter 9). In allowing for the proactive engagement of Indigenous peoples in research, every element of these protocols and principles sanctioned under the RHS and OCAP incorporate aspects of data sovereignty to some degree (Napoleon, 2015). Some of the protocols are initiated by Indigenous peoples on their own as well as by the research funding agencies themselves or professional associations as a matter of best practices (Bannister, Chapter 12; Burelli, Chapter 13). As the drive for big and open data continues to define the future of research, IPLCs' push for data sovereignty will find stronger and wider traction for critical data analysis to mainstream the imperative for data equity as a moderating principle despite all the benefits canvassed for big and open data phenomena.

CONCLUSION

Fueled by ICTs, big and open data phenomena constitute one of the most progressive infrastructures for the advancement of research in the 21st century. Big and open data reduce costs, enhance the democratization of vital research data and related information, including those involving GRs, aTK and, broadly, research relating to IPLCs. Yet the convergence of bio-digital technologies and more broadly the paraphernalia of the ICTs in GRs and aTK contexts practically results in the de-linking of critical Indigenous related research data from their origins and sources in Indigenous and local communities and their knowledge systems. These technological realities are real grey areas exposing gaps in the legal and policy framework for ABS under the NP, CBD and cognate regimes. They also remain a significant source of present and future challenge to implementation of ABS. As argued by Phillips, Smyth and de Beer (Chapter 10), not only did the NP fail to 'deal with the pressing issue of digital

technology transfer,' the Protocol was already rendered obsolete by technological developments before it came into effect. Counterintuitively, however, a critical analysis of aspects of the relevant provisions of the Protocol and cognate texts of relevant international instruments on the utilization of GRs, aTKs and the concept of derivatives support the *continued* relevance and accommodation for equitable ABS in era of technology's tendency to conflate sources and origins of GRs and aTK. Even if that is not enough, a critical appraisal of progressive developments on the rights of IPLC generally reinforce the relevance of ABS implementation through ongoing technology transformations as a matter of justice and equity.

Rather than undermine ABS, overall, the current state of affairs energizes momentum for pragmatic and responsive incorporation of critical data studies and data equity to big and open data phenomena as an important site for progressive policy making toward an Indigenous-sensitive ABS. Using critical data analytical approaches and constructive deployment of data sovereignty, IPLCs are better able to sustain their demand for equitable ABS in the wake of new technologies. To this end, the novel concept of data sovereignty is one that should command attention for policy elaboration and for balancing converging interest in Indigenous research, ABS and the big data and open data phenomena. As Oguamanam and Hunka pointed out in Chapter 3, the use of technologies in ABS-related contexts is one of the priority areas of Indigenous capacity building and capacity development for equitable ABS in the Canadian context.

REFERENCES

Amani, Bita and Rosemary J. Coombe. 'The Human Genome Diversity Project: The Politics of Patents at the Intersection of Race, Religion and Research Ethics' *Law and Policy* (2005) 27:1 152–88.

Bagley, Margo and Arti K. Rai. 'The Nagoya Protocol and Synthetic Biology Research: A Look at the Potential Impacts.' *Virginia Public Law and Legal Theory Research Paper* (2014). No. 2014–05. Online https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2388794.

Briggs, Eileen, Patricia Riggs, and Nancy Palmanteer-Holder. 'Data as a Strategic Resource: Self-Determination, Governance, and the Data Challenge for Indigenous Nations in the United States,' *International Indigenous Policy Journal* (2017) 8:2.

De Beer, Jeremy. 'Ownership of Open Data: Governance Options for Agriculture and Nutrition' (Wallingford: GODAN, 2016).

Hilbert, Martin and López, Priscila. 2011. 'The World's Technological Capacity to Store, Communicate, and Compute Information.' *Science* 332 (6025): 60–5.

International Indigenous Data Sovereignty Interest Group, International Indigenous Data Sovereignty IG Charter Statement. 2017. online www.rd-alliance.org/group/international-indigenous-data-sovereignty-ig/case-statement/international-indigenous-data.

The United States Indigenous Data Sovereignty Network (USIDSN) Conference Report, Enabling the Data Revolution: An International Open Data Roadmap, 3rd International Open Data Conference 2015, online: http://1a9vrva76sx19qtvg1ddvt6f.wpengine.netdna-cdn.com/wp-content/uploads/2015/09/IODC2015-Final-Report-web.pdf.

Meuller, Milton. 'Summary.' *Will the Internet Fragment?: Sovereignty, Globalization and Cyberspace* (Polity Press, Cambridge: 2017).
Oguamanam, Chidi and Vipal Jain. 'Access and Benefit Sharing, Canadian Aboriginal Research Ethics Policy after the Naogoya Protocol: Digital DNA and Transformations in Biotechnology' (2017) 3:1 *Journal of Environmental Law and Practice* 79–112.
Scassa, Teresa and Frazer Taylor. 'Legal and Ethical Issues around Incorporating Traditional Knowledge in Polar Data Infrastructures' (2017) 16 *Data Science Journal* 16:1–14.
Schultz, Jennifer and Stephanie Rainie. 'The Strategic Power of Data: A Key Aspect of Sovereignty,' *International Indigenous Policy Journal* (2014) 4:5.
Spector, Tim. 'Three Days as Hunter Gatherer.' The Conversation June 30, 2017, online https://theconversation.com/i-spent-three-days-as-a-hunter-gatherer-to-see-if-it-would-improve-my-gut-health-78773.

NOTES

1 Emphasis added.
2 WIPO/GRTKF/IC/35/4.
3 There is no definition of research in any of the instruments, NP and the CBD.
4 Arguably, this is a basis to suggest that the CBD is not oblivious of digital sequence information (DSI) in relation to TK associated with GRs.
5 Carleton University, Geomatics and Cartographic Research Centre and Canadian Internet Policy and Public Interest Clinic (CIPPIC), *A Proposal: An Open Licensing Scheme for Traditional Knowledge* (2016), online: https://cippic.ca/sites/default/files/file/CIPPIC_GCRC–TK_License_Proposal–July_2016.pdf.
6 University of Victoria.
7 (www.thememo.com/2017/07/12/hadza-gut-health-improvement-microbe-diversity/) Microbiome project with Hadza Indigenous Hunter Gatherers of Tanzania who are linked to 50, 000 generations of human ancestry . . . linked to cradle of humanity in Africa . . .).
8 Fridtjof Nansen Institute, 'The Concept of "Genetic Resources" in the Convention on Biological Diversity and How it Relates to a Functional International Regime on Access and Benefit Sharing' (2010), online: www.cbd.intdocmeetingsabsabswg-09information abswg-09-inf-01-en.pdf.