

PETER CLEAVE

Maramara me te Iwi Taketake: Data and the Indigenous Group

Abstract

This paper considers Indigenous groups and data. The paper begins with fifteen assorted questions which are addressed in various ways in the next two sections. The second section is a review of 'Indigenous Data Sovereignty' a collection by Kukutai and Taylor of 2016. This collection is seen as an excellent statement of the position of the Indigenous group regarding data and each chapter is reviewed in several paragraphs. Beginning with Kukutai and Taylor, the third and final section is a commentary on recent literature on data with reference to the Nation-state, Big Tech and Indigenous groups. This section considers a shifting situation involving machine learning and the hunting, gathering and farming of data. A reappraisal of the way data is used in the context of the Indigenous group, the Nation state and Big Tech

Professor Peter Cleave is a Research Associate in the Faculty of Health and Environmental Sciences (AUT University), a Visiting Fellow in Te Ara Poutama (AUT University), a Member of Common Room, Wolfson College (Oxford University), and a Broadcaster in Te Reo Maori at Kia Ora FM

is proposed. That reappraisal involves new considerations of identity in forms of ethnicity, nationalism and tribalism as well as the way Indigenous groups are defined by others and the ways in which they define themselves.

Assorted questions

1. How does data define the ethnic group?
2. Who owns what data about whom? Why?
3. Data feeds profiles?
4. Data is a mirror?
5. Data is resource?
6. Data is money?
7. In data we trust?
8. Data feeds the future?
9. Data is identity?
10. Does Machine Learning throw up a field of data that imprisons the imagination?

11. Do data and Machine Learning provide a field of dreams?
12. What are the critical criteria for data collection by Indigenous groups, Nation-states or Big Tech?
13. We are known by our data?
14. Does data define the Nation-state?
15. Data is power?

Review of *Indigenous Data Sovereignty, Towards an Agenda* (Kukutai & Taylor, 2016)

Before colonisation data was controlled from within Indigenous groups and not from without. After colonisation and in a variety of ways Indigenous peoples came to be dominated by nation states. One of those ways was the taking control of the assemblage of data, its processes of collection, the way it was and is structured, its storage and its use. Kukutai and Taylor as Editors and all contributors to the collection that is *Indigenous Data Sovereignty, Towards an Agenda* comment on data control in various ways.

Indigenous Data Sovereignty opens with an Introduction by Kukutai and Taylor entitled *Data Sovereignty for Indigenous Peoples: current practices and future needs*. For anyone interested in that kind of a thing there is a several page list of acronyms at the outset of the book and this is good because there is a bewildering amount. After this and a useful Preface the Introduction begins with reference to the CANZUS conference in Canberra in 2015. The idea was to ‘identify and develop an indigenous data sovereignty agenda, leveraging international instruments such as the United Nations Declaration on the Rights of Indigenous Peoples (UNDRIP).’ Kukutai and Taylor make the claim which seems accurate that this is the first book to ‘engage with the topic of data sovereignty from an Indigenous standpoint.’

The thrust of the collection is to shift from ‘servicing government requirements’ to support ‘indigenous peoples’ development agendas’, to shift from a development-policy nexus in nation-state settings and to reflect on how the statistical portrayal of Indigenous peoples might be transformed. Various points about census taking are considered to emphasise this thrust.

Kukutai and Taylor note that in the present situation for indigenous peoples around the world there is a tendency to generate ‘crude social binaries (indigenous-non indigenous) as input to public policy’. They conclude the early part of their paper saying, ‘...while not denying some role for centralised data collection, what indigenous peoples are seeking is a right to identity and meaningful participation in decisions affecting the collection, dissemination and stewardship of all data that are collected about them.’

The above might be taken as the central premise in the collection.

Megan Davis in Chapter Two introduces the topic of the ‘Indigenous Navigator’ project, a way of reporting on the implementation of Indigenous rights in development and other matters. Her paper is entitled ‘Data and the United Nations Declaration on the Rights of Indigenous Peoples’. Davis starts by stating the invisibility of indigenous people in data collection of national statistics. Citing Stewart-Harawira (2005) and others Megan Davis asks

whether this is ‘to avoid considerations of indigenous peoples’ right to self determination’ and goes on to set out questions relevant to data sovereignty. Davis points out logistical challenges and gaps in the case of data collected about indigenous peoples and mentions the fact that many Indigenous peoples work in informal economies so data collection is difficult in such areas. She asks how data might be dis-aggregated so that indigenous data might be more clearly seen. Davis mentions the ‘indigenous identifier’ used in some Latin American studies.

Davis then discusses UNDRIP and data suggesting that it provides ‘the common framework of the normative content of the rights of indigenous peoples and is therefore important to understanding the issues of indigenous data sovereignty’. There is then a good discussion of UNDRIP followed by a consideration of how to measure the implementation of the human rights standards that it affirms be this by ‘fact based’, objective or ‘judgement based’ subjective methods. Davis concludes with a discussion of the Indigenous Navigator project now found in several UN organisations.

Mathew Snipp in the next chapter examines the concept of ‘data sovereignty’ and asks how indigenous people might claim and gain greater control over data related to them including determining access to that data. As Snipp mentions

at the outset, data sovereignty ‘is a uniquely twenty-first-century expression that arises directly from the explosive growth of information associated with the internet and the spread of mobile phone technology’. Regarding the role of the Nation-state in all this he gives the example of Google leaving China to be replaced by Baidu.

Snipp asks from whom might data be obtained and talks about the way indigeneity might be defined and offers 1492 as a benchmark in the Western Hemisphere defined in various ways by different nation states. In Mexico speaking an Indigenous language is a key indicator. Various points are made about defining tribes and recognising them in the USA including the idea of blood quantum.

Some people report that they are American Indians but do not state a tribe (20%). This makes it difficult for tribes to assess data on their members. They do not have the means to do this.

Then there is the question as to where people reside, many living across cities rather than in discrete areas.

Snipp goes on to ask what is data about and what is it for? Surveillance by a settler state or to assess the specific needs of the indigenous community? He suggests that African Americans, Asians and Hispanics are better treated regarding data. He notes that American Indians and Alaska Natives make up only 2% of the total US population.

Snipp proceeds to consider privacy issues and mentions blood samples in a case from Arizona where data was widely shared. There are issues of scale, the Chapter Houses of the Navaho being too small to be considered by the Census Bureau, He talks about the use of mobile phones and suggests that they might make things easier in data collection by indigenous groups. He concludes with several points including that Indian groups themselves define their membership and who has access to the data collected.

Ian Pool then writes about the pre and post colonial experience for indigenous peoples and notes the entry of 'big data' to this situation. He offers the following definition, 'Data sovereignty (Dsov) is a somewhat narrow twenty first century concept from commercial law relating to the protection of digitalised individual, governmental and corporate information, and also to the safeguarding of the national security apparatus from nefarious actions.' Pool, in an article that covers a fairly wide field talks about 'parastatal' corporations like the East India Company in its day in terms of what he calls Data suzerainty, D Suz, and compared to Data sovereignty, D Sov. This distinction is useful and it is worth thinking about these parastatal corporations and Big Tech.

Pool talks about external data systems. He walks the reader through a consideration of colonialism and talks about the way settler states demonised

indigenous groups and subverted their datasets. Pool argues that what was left such as songs and dances 'constitute a form of memory bank for data'. He then provides 'a backstory' to show the kinds of effects involved which include 'methodological chasms between indigenous peoples and the wider populations surrounding them'.

In a fairly wide ranging discussion involving cultural capital, the degradation of the native, Tabloidism and the prioritisation of certain statistics like those relating to morbidity *inter alia* Pool goes on to internal colonialism. He suggests that people might lose data sovereignty when their credit card details are held offshore. He mentions data hoovering and talks about classification. Pool finishes by saying 'It would be tragic if this metamorphosed instead into neo-D suz under transnational corporate rule beyond the control of indigenous peoples or the polity in which they live.'

Maggie Walker then talks about the five 'D's, of indigenous Australian data- disparity, deprivation, disadvantage dysfunction and difference. Then there is the idea of 'deficit data-problematic people' and statistically informed pejorative stereotypes. Walker talks about a 'recognition space'. Maggie Walter asks about the reality of numbers and their deployment and their service throughout her paper. Then it's a matter of 'Closing the Gap', a pervasive sense of data inequality for a group, a weighting of ethnicity. Is this a closing of the gap between one

sense of group and another? In the way nation states collect data this is sometimes about scale regarding the small size of dispersed populations. Walter talks about ‘the trope of the statistical Indigene’ outside of which there is, she argues, a data desert’.

But the 5Ds add up to a frame which structures race and other relations. And provide a ‘virtual veil’ over the situation of indigenous peoples. The frame allows a new racism. Walter comes to say,

‘The advent of big data with its tendency to further distance lived social and cultural realities from their database embodiment, has only exacerbated the pejorative power of numbers to further marginalise and dispossess.’

Walter asks how the paradigm of indigenous statistics might be disrupted and disturbed. Later in a discussion relating to Larrakia sovereignty Walter also discusses how to disrupt established tropes regarding indigenous identities. As she concludes Walter talks about a world view of those who shape how indigenous statistics are ‘done’ referring to Bourdieu’s 1984 idea of ‘synthetic unity’ she does so. Finally Walter argues for ‘Indigenous framed numbers’.

Francis Morphy then talks about demography and argues that local understandings are important and talks about ‘naming’ the indicators of indigenous realities and prioritising those indicators. Morphy talks about ‘enforced commensurability’ and how

numbers are use against indigenous people. She talks about ‘the ontology of the quantifiable’ and gives examples of data being used in a negative way for indigenous people. Morphy talks about two major aspects of data sovereignty. ‘It is not just a matter of contesting decisions based on indicators preordained by others; it also involves the assertion of sovereignty over the choice ODF indicators’. Morphy argues for indigenous peoples to have ‘epistemological sovereignty over the data that define them’. There is data and the way it is aggregated. The way indigenous people are compared and ranked shows the ‘enforced commensurability’ that they live under. Referring to Davis *et al* (2012) Morphy talks about how indicators implicitly evaluate. There is an interesting and possibly important discussion of place which Morphy considers critical for indigenous peoples.

Morphy talks about the ways in which indigenous demography emphasises ‘substantive differences that need to be acknowledged and accepted by settler states if they are to formulate policy that supports rather than undermines the self defined goals of encapsulated indigenous peoples.’ Morphy closes with a consideration of polygyny.

Diane Smith then asks who is the self in ‘self-determine-nation;? Smith argues for ‘culture smart’ information’. Referring first to a quote from Einstein ‘(according to the available data)’ worth considering,

‘Not everything that can be counted counts. And not everything that counts can be counted’.

Smith looks to the ‘internal expertise and institutions’ required by indigenous groups regarding data along with access to data that is culturally relevant and accurate. She lists a series of questions that might yield the collective will of a nation, clan, group or community’ in such a way as to ‘mobilise sustained, organised action’. Smith is talking about the governance of data and the abuse that occurs from ethnocentric criteria and definitions, about the way a ‘*nullus* fiction’ of indigenous life occurs to leave it ‘empty, invisible and unknowable’. At the same time there are challenges to indigenous data governance. There is a need now to ensure ‘resilient governance capable of delivering outcomes’.

Diane Smith then provides a useful model with several dimensions requiring data and information. She asks where to start in matters of data sovereignty and asks searching questions as to the ‘self’ involved, the appropriate demography and governance performance. Smith talks of a ‘virtuous’ (her inverted commas) cycle of relevance and goes on in a section entitled ‘Culture-smart information’ to speak of a ‘tyranny of the measurable’ which does not meet the ‘complex knowledge economy’ found in indigenous society. She sees ‘culture-smart’ data as ‘fundamental ingredients in the practical exercise of sovereignty’. In her conclusion Smith goes on to say,

‘Sovereignty includes being able to design rules for the restriction and opening of data’.

Ceal Tournier then describes a way of certification from an indigenous perspective. He speaks of a shift after 1994 in ‘data and information sovereignty’. To Tournier ‘the problems with the use of First Nations’ information stem from who is in control- and thus what gets done, how it is done and who knows about it’. Ceal Tournier, as Chair of the First Nations Indigenous Governance Centre (FNIGC), said at an earlier point, ‘he who controls the data controls the gold’ (Tournier 2002).

Tournier goes on to say, First Nations’ citizens and leaders acknowledge and act on the premise that information needs defending and protecting; just as we protect our lands, our forests, our animals and our fish, we need to protect our data which are an extremely valuable renewable resource’.

Tournier then gives examples of poor research by outsiders involving data.

Tournier proceeds to talk about a *red standard* (Tournier’s italics) emphasising First Nations control of First Nations’ data and provides an important history of the experience in Canada involving Ownership, Control, Access and Possession. Tournier points out that those responsible must be versed in Jurisdiction, Policies and Procedures and Repatriation. He suggests that while these apply to Canada they may, with modification, be applied worldwide by indigenous

peoples worldwide. He closes by saying, 'By building information governance capacity, enacting our own laws, entering into data-sharing and license-to-use contracts, creating regional data centres and repatriating our data First Nations are getting closer to exercising full jurisdiction over our information'. Then Hudson, Farrar and McLean talk about the Whakatohea and ask for equality of access to data and a datascape where rights to culturally sensitive data be worked out fairly for iwi. The example given first is the Whakatohea Maori Trust Board as 'the steward for a range of administrative datasets, health and social service records, commercial information, historical accounts, indigenous knowledge, strategy documents and research'. They talk of co-governance, this happening in the post settlement period. The Whai Rawa superannuation scheme of Ngai Tahu and the business incubator model of Waikato-Tainui are given as development examples in the post settlement period. The scale of the Maori economy is examined as well as the growth of Maori intellectual capital. They note that, 'Through the increasing range of co-governance and co-management relationships tribes are asserting the importance of including indigenous knowledge as an information source in environmental decision-making.

Regarding iwi and data Hudson, Farrar and McLean suggest, 'Using a health analogy it could be said that iwi suffer from inequities in data access and

inequalities in data infrastructure'. They then go on to talk of the 'five safes' framework used in the Integrated Data Structure (IDI), a 'linked longitudinal dataset'. These include safe people, safe projects, safe settings, safe data and safe output. Referring back to the New Zealand Data Futures Forum of 2015 they give four principles that include Value, Inclusion, Trust and Control. In a useful discussion of tissues and genome data they refer to the Ngati Porou Hauora agreement with the University of Otago on the genetics of Gout and other matters. The discussion proceeds to the Mataatua Declaration on the Cultural and intellectual Rights of Indigenous Peoples of 1997 as reinforced by the United Nations Declaration on the Rights of Indigenous Peoples (UNDRIP). There is then a discussion of research ethics informed by their earlier points. This is followed by a review of resource rights and interests and the use of the Waitangi Tribunal in settlements which as lead to the creation of pan-tribal entities such as Te Ohu Kai Moana or the Crown Forestry Rental Trust..

The Whakatohea Maori Trust Board is then used as a model for the use of data by an iwi such as the creation of critical data archives. Hudson, Farrar and McLean go on to say, 'Whakatohea consider that access to government collected data and information is a treaty right and that the tribal entity is better placed to create benefit for tribal members than government departments or research

organisations’. They then provide a model with exclusive rights, shared rights and shared interests. James Hudson then talks about the Maori Plan in Tamaki Makau Rau, Auckland, He suggests that , ‘The Maori Plan clearly articulates a Maori vision for Tamaki Makaurau and is a touchstone for ongoing dialogue about Maori-specific data to support that vision’.Regarding the Treaty of Waitangi and local government Hudson talks of partnership, active protection and redress. He then talks about a duty to consult. Following a discussion of governance and operational matters to do with the Independent Maori Statutory Board Hudson considers the Maori Plan. He says, ‘While some data were available *about* Maori, few, if any, were available that could be viewed as useful *for* Maori’ (Hudson’s italics) and proceeds to discuss a framework for monitoring and measuring Maori outcomes in Tamaki Makau Rau. The Maori Plan is to follow tikanga, last for thirty years and address a broad range of issues. Hudson mentions the difficulty in sorting out national from local data especially in environmental and cultural factors. He considers the establishment of a Data Strategy Expert Panel and notes an emphasis on ‘safety’ (Hudson’s inverted commas) around tikanga Maori and ‘technical robustness’.

Rawiri Jansen then talks about an indigenous professional class in Aotearoa New Zealand and shows how data can be mobilised to inform action by Maori. He does this through case studies showing

'how data can be used to inform interventions that the NHC leads to deliver improved health outcomes for Maori'. Going back to Nopera Pana-kareao he suggests that 'Data governance requires thought about both the substance and the shadow of the data'. He talks of health rights and the need to 'deconstruct institutionalised racism'. Taking rheumatic fever and the Mana Kidz programme as examples Jansen points to the value of linked data in providing a 'rich picture'. Then he talks of housing plans and 'housing literacy' and the construction of a dataset to better manage rheumatic fever. He describes the use of an AWHI scorecard to follow developments. Jansen goes on to discuss cardiovascular disease and use of Mohio Forms available on the internet to monitor data and provide a 'rich data picture' which helps in measuring equity performance.

Jansen talks about sharing data and mentions the Metro Auckland Data Stewardship Group (MADSG) 'to navigate the issues of health information use, management and privacy across the health system'. He talks about the idea of a Maori 'data-sharing governance framework' which would be 'an indigenous data sovereignty framework'. He concludes by saying, 'It seems that we move up a hierarchy from data visibility and data accessibility to data sharing and data control'. Jansen's final sentence is 'Maori sovereignty is informed by Maori data sovereignty'.

Ray Lovett then talks about the need for professionally trained indigenous statisticians in the Aboriginal and Torres Islander community. He says the ‘need to reflect reality is what Walter and Anderson (2013) refer to as ‘the cultural framework of Indigenous statistics’. Lovett talks of local distrust in data research and then the need for Indigenous statistical capacity. He points to suggestions that the reliability of Aboriginal and Torres Strait Islander data has been negatively affected by ‘the purpose of the data collection, who is conducting it and the mode of collection’. As well as issues of capacity there are matters to do with the chairmanship of committees and groups that ‘advise (Lovett’s inverted commas) and not direct. He goes on to say that while ‘the gap’ (Lovett’s inverted commas) is not closing, important cultural strengths are not noted. He asks, ‘What of the distal indicator of a strong connection to mob and country showing a positive correlation with reductions in cardiovascular disease outcomes? (Rowley *et al.*2008)’ and goes on to talk about indigenous conceptions of well being and the value of Indigenous culture. Lovett discusses the teaching and learning of mathematics which is critical in developing statistical capacity and laments the financial costs of higher education. He mentions a programme called Mayi Kuwayu and the development of statistical capacity in longitudinal

studies and concludes by emphasising the need to build statistical capacity.

Yap and Yu then talk about indigenous data sovereignty in the case of the Yawuru. A Yawuru language revitalisation program is included in the discussion. They begin by addressing the flaws they see in the ‘Closing the Gaps’ policies. This is not all about adopting the ways of the majority. They refer to Taylor’s 2008 suggestion of a ‘recognition space’ and then to Kukutai and Taylor’s 2013 insights into how data might be indigenised. They argue for the development of ‘indicators that are culturally appropriate and reflect indigenous world views’. They call for ‘information on cultural dimensions, indigenous ecological values and indigenous peoples’ unique relationship to nature and the living landscape.’

Yap and Yu ask ‘how does one begin to operationalise the recognition space so that the information reflects Indigenous aspirations and world views while simultaneously informing government planning and reporting needs?’ They refer to five recognition principles developed by Kukutai and Walter 2015. These are geographical diversity, cultural diversity, other ways of knowing, mutual capability building and indigenous decision-making.

Yap and Yu proceed to an analysis of Aboriginal and Torres Straits Islanders beginning with the Native Title Act of 1993 which provided a legal framework

for a self determination agenda. They suggest ‘The Native title process provided a platform for the process of knowledge building and capacity building among Yawuru.’ They then describe the Yawuru Knowledge and Wellbeing Project which involves mapping and other features. The case studies described deserve to be considered in detail and a thorough reading of their work is recommended. The study seems highly appropriate as, to use the words of the authors, ‘the reality of native title sets in’.

Desi Rogriguez-Lonebear gives examples from the USA regarding data identity and tribal membership. He points out that in the ‘data mecca of the United States American Indian tribes face a paucity of data about their own populations.’ He explores ‘the nascent data revolution now getting underway in Indian country as some tribes regain data sovereignty’ and considers ‘how tribal data sovereignty can be a powerful tool in decolonisation and in pursuit of tribal developmental goals.’

Rodriguez-Lonebear notes that ‘For indigenous peoples data were everywhere and survival was often tied to one’s ability to gather, analyse and share this knowledge’. He also notes the importance of oral history before considering how data sovereignty became data dependence with a recounting of what happened in a museum context with the brain of an elder.

Roderiguez-Lonebear argues ‘Indigenous data engagement in the United States is inextricably tied

to the subjugation of American Indians and federal policies of Indian extermination and assimilation.’ He talks about the omission of the majority of American Indians from the census for a very long time because they fell under the ‘Indians not taxed’ category. He suggests that ‘Indigenous data transitioned from a means of survival to a mechanism of federal administration’. He argues, ‘The data collection activities of tribes now largely revolve around mandatory federal grant recording and many tribes employ grant officers or administrators to oversee these efforts.’ He draws a distinction between New Zealand and the USA and points to the difficulty tribes face in gathering data about themselves. This makes policy making difficult.

Rodriguez-Lonebear goes on to argue that this is why often the data tribes themselves present is regarded as inferior to that collected by the state governments such as that of Montana. He talks of tribal youth being the ‘data warriors’ needed to steer a data revolution. He raises the question as to who should be conducting censuses, the tribes or the US government. He notes the difficulty in maintaining sovereignty in a dependent situation and suggests that data sovereignty is critical. He also notes debate on the idea of blood quantum and the need to have good data on tribal membership. He suggests in conclusion that the silo mentality must be dropped

and that cooperation is necessary for data sovereignty.

Paul Jelfs then discusses the Australian Bureau of Statistics in terms of indigenous governance in the case of Aboriginal and Torres Strait Islands. Jelfs gives a strong paper on the role of the ABS in this area talking about such initiatives as the Indigenous Community Engagement Strategy, the Roundtable on Aboriginal and Torres Strait Islander Statistics and other things such as Census Story Books. He concludes with a model for ‘solution-centred and collection-centred statistics’.

Finally Darin Bishop talks about the Maori Statistics Framework and the conceptualisation of Maori indicators. He mentions systems found around the world regarding statistics of indigenous peoples. Bishop suggests that ‘New Zealand has gone further than any other nation-state in seeking to develop such systems and to accommodate the needs of the indigenous people. Notwithstanding this, significant deficiencies remain and these undermine Maori data sovereignty.’ Bishop provides a model of the measurement process and also a critique and the need for dis-aggregation of data and to follow recommendations from the United Nations Permanent Forum on Indigenous Issues.

He talks about the shift away from ‘closing the gaps’ to a focus on Maori potential. Bishop speaks of the need to provide Te Puni Kokiri, the Ministry of Maori Development established under the Ministry of

Maori Development Act of 1991, with ‘relevant and reliable evidence’. Bishop notes the small window of opportunity for Maori to act on data and proceeds to offer a set of suggestions for users and potential users of Maori statistics. He mentions Te Kupenga of 2013 a survey of Maori wellbeing following a census. He talks about limitations on data pertaining to whanaungatanga.

There is also a discussion of an emphasis on the individual as distinct from notions of the collective when it comes to ethnicity. Bishop proceeds to list a series of challenges ranging from Maori migration patterns regarding New Zealand and Australia, to cultural outcomes and various other important areas and argues for the voice of Te Mana Raraunga, the Maori Data Sovereignty Network. In conclusion Bishop goes back twenty years to look at the development of a Maori statistics framework and reviews progress along with deficiencies that remain. At the close of an excellent discussion he mentions the scope for building the statistical capability of Maori communities and organisations amongst other things.

Going to the voice of Te Mana Rarauna the Maori Data Sovereignty Network the following points are made:

1. asserting Maori rights and interests in relation to data

2. ensuring data for and about Maori can be safeguarded and protected
3. requiring the quality and integrity of Maori data and their collection
4. advocating for Maori involvement in the governance of data repositories
5. supporting the development of Maori data infrastructure and security systems
6. supporting the development of sustainable Maori digital businesses and innovations

Conclusion

We have a collection of some note in *Indigenous Data Sovereignty*. Kukutai and Taylor conclude their excellent introduction by referring to UNDRIP as they have throughout. They note the distinction between 'sovereignty as it relates to digital spaces and the forms of data stored in those spaces'.

Differences in the CANZUS states are noted. The United States Indigenous Data Sovereignty Network and its four focus areas of data for sovereignty, data collection and access, data storage and security and data as intellectual property are pointed out.

Kukutai and Taylor stress the emphasis that state agencies have placed on 'populations' rather than 'peoples'. They point out that they have provided a voice for informed people about indigenous sovereignty. They ask, again with reference to UNDRIP, for 'a relocation of authority over relevant

information from nation states back to indigenous peoples’.

Comment

In describing a displacement of the situation whereby the Indigenous entity controls data to one where the nation state controls the data Kukutai and Taylor and most of the contributors to this volume do not fully round on the role of ‘Big Tech’. In the introduction Kukutai and Taylor do summarise Ian Pool’s argument and talk about the irony Pool finds when just as the indigenous peoples are coming to terms with the data that the Nation-states have of them ‘Big Data’ is emerging to engage in and in some ways own the data of the Nation-states. This, by the present writer’s extension, and not necessarily with reference to Kukutai, Taylor or Pool, might mean a progression and a replication whereby the nation states get left in the position that they had created for indigenous minorities or something like. As the Nation-states in relation to Big Tech or what the present writer calls the ‘algorithmic state’ (Cleave 2020, 2021) begin to look like the Indigenous groups in Canada, the US, Australia, New Zealand and elsewhere in relation to those Nation-states. The Nation -states, especially the smaller ones, are like stars that are fading, at least for the moment, while the stars of Big Tech are waxing.

Another way to put this is that in making the comparison we might be moving from one 'uncanny valley' to another one just as 'creepy' to borrow terms from Sara Watson (2014) in her discussion of what she calls 'data doppelgangers'. Watson gives a background to her terms,

'Since the 1970s, theorists have used the term "uncanny valley" to describe the unsettling feeling some technology gives us. Japanese roboticist Masahiro Mori first suggested that we are willing to tolerate robots mimicking human behaviors and physical characteristics only up to a point: When a robot looks human but still clearly isn't.

The threshold is where we shift from judging a robot as a robot and instead hold it against human standards. Researchers at the University of Bolton in the UK have described this shift as the "Uncanny Wall" in the field of digital animation where increasing realism and technological advancements alter our expectations of how life-like technologies should be. I would argue that we hit that wall when we can't distinguish whether something is broadly or very personally targeted to us. The promise of Big Data has built up our expectations for precise messaging, yet much of advertising is nowhere near refined. So we don't know how to judge what we are seeing because we don't know what standard to hold it against.' (Watson, 2014).

Sarah Watson's expression 'data doppelgangers' refers to the situation where a false sense of the

individual is created, one that the individual has no control over the shape or shaping involved. The same might be said of the Indigenous group as represented in the data collection of the Nation-state. To return to Pool's point, the emerging role of Big Data takes this to a further remove in that the representation of Nation states themselves as well as the ethnic groups within them, Indigenous and otherwise, come, in a collective rather than an individual way, to make up a set of data doppelgangers. Considering Watson's argument, what then the relation of the indigenous entity to Big Tech or what the present writer calls 'the algorithmic state'? And what if, referring back to Maggie Walter's points about data deprivation above, in the absence of a positive relationship between Big Tech and Indigenous groups the latter simply stay in data deserts? What if, instead of uncanny walls or valleys there is an eerie silence upon a plain. What if ethnic groups are simply left off the map as the hunting, gathering and farming of data proceeds? What if, when we look at such a high and low tech contrast we are looking at a version of the Hot and Cold societies described by Levi Strauss between history based and myth based societies (Levi Strauss 1966)? And if we closely looked at this would we see that while some may be data poor they might be rich in myth, wealthy in an idea of themselves that makes sense to the group and offers emotional and psychological support?

There is a general question that might apply. Is the machine learning found in Big Tech set to use algorithms to know you as a citizen or to know you as a consumer? To turn on this point, as the Nation-state to the Indigenous group is replicated in the relation of Big Tech to the Nation-state the nature of the knowledge involved changes. No longer are we talking about data and rights and obligations of a civic kind, at least not necessarily. In its most vulgar form the data context of Big Tech and the individual member of a Nation-state or an Indigenous group is basically a market context. A market economy without reference, say, to a political economy or, for that matter, to a political identity.

Whose data context, whose economy is this?

Now the nation state is being displaced in certain respects by the algorithmic state or at least the present writer suggests so (Cleave 2020 One- Four, 2021 Five-Eight). This especially relates to data. 'In data we trust', says an Amazon executive (Cleave One *op cit*). But, again, who owns the algorithmic state and how is it directed? The first part of this question is, at least at first glance, easily answered. Amazon, Alibaba and the like and their shareholders own Big Tech. Coming to the second question, as to how all of this is directed, we come to machine learning. Again the reference here might be to *The Algorithmic State History and Theory* (Cleave, Five *op cit*.) and the idea that it began with Bayes. The point might be that this is a direction by algorithm and

machine learning for a great if not the most part and this might be new in human history even though the historical precedents stretch back to the publication of Bayes' work (Bayes 1763).

To consider the above propositions several examples, albeit from the silly side of the sanity ledger might be considered. The media in Myanmar is dominated by Facebook which will not allow commentary by the government, said government being the military. San Salvador is considering allowing and encouraging Bitcoin as a valid currency in the context of a weak local currency. In 2021 there has been the example of Twitter and other platforms banning the President of the United States from commentary. In these examples cardinal principles of state may be seen compromised, communications in the Myanmar case, economy in the San Salvador example and politics in the case of Trump as President of the United States of America. At the same time Big Tech is looking strong against competing interests in the fields concerned, media, economics and politics.

If we were to find a collapsed or failed nation state that allowed Facebook and Bitcoin which also gave free rein to Twitter and the like we might be looking at a displacement of some significance.

Throw in neighbourhood meshes and the like...

It might be worth pointing out that in Aotearoa-New Zealand of the early 2000s this is the time of 'He Puapua'. This is an attempt by Willie Jackson and

other Maori MPs to provide more say in government departments for Maori. But what is the value of having more say in areas where data ownership and control might be decreasing? Is the effort to create 'He Puapua' a chasing of shadows, phantoms of power? If data control is increasingly in the hands of Big Tech are we better to ask the relation between Indigenous groups and Big Tech than that between Indigenous groups and Nation states? We might consider the review of Kukutai and Taylor above in these terms.

Will Nation states rein in Big Tech and attempt to remain or become again the key players in data sovereignty? There is a serious situation here for Indigenous groups. If Indigenous groups are trying to gain control of data, to find again their sovereignty of data in the wrong place then there is a twist again to colonialism. This is a sovereignty game of some portent in that the Kukutai and Taylor collection shows the kind of energy going in to claim a resource that may not be there in the Nation-state, a resource that has shifted in the time of the data claiming process, over the last quarter century, say. Is there a grip that the indigenous society might secure on the algorithmic state on 'Big Tech', a hold like UNDRIP might have given indigenous interests over the nation state?

What is the significance of the data shift, the shift from state to parastatal or what the present writer has called the algorithmic state or, more simply, the

shift to Big Tech as the critical owners of data? Given the points about identity and sovereignty above the shift is considerable. There is a shift in shapes, political, epistemological and other as well as a shaping of ethnicity and nationalism.

Is this some kind of migration from one data zone to another, one holding pen of information to another? Is this, at base, all about data as clicks? Every movement of the mouse and every dwelling upon or return to a page or a link is in the zone of Big Tech rather than in the data zone of the nation state. And each click is a point in the data bank, the holding pen of data of, say Google or another Big Tech operator.

There is a mix and sometimes a contrast between skills and sentiment in the Kukutai and Taylor collection as there is in a lot of the commentary about data. The 'skills' relate to the algorithms involved, the efficiency and the effectiveness of the data collection involved, the skills in the hunting, fathering and farming of data. The sentiment often relates to the Nation-state or, throughout this collection, the Indigenous group involved, to the critical identity point upon which to hang data.

There are also issues of scale, the Indigenous groups generally being too small to allow effective, cost effective data collection. As mentioned above Morphy states very well the problems of small scale data collection for Indigenous groups. But the same

logic might apply for the small nation state like New Zealand.

With the situation of the very small nation state and the bigness of Big Tech to consider there is also a strangeness of place, such places as New Zealand being easily seen and identified by virtue of their isolation from continental Europe or America. At the same time there is a placelessness of Big Tech, the latter existing in several spaces including cyberspace without being defined by the bounds of the Nation-state.

With all of this there is also, as introduced above with the consideration of sentiment, what might be called a constellation of comfort, a cluster of sentiments. This comfort constellation, this sentimental structuring of sovereignty is seen throughout the Kukutai and Tatlor collection with, say, Pool's curation of the New Zealand context. The cultural and political loading in Pool's argument functions as a kind of algorithm following Cathy O'Neill's use of the term (O'Neill 2018). O'Neill uses the example of a recipe as a kind of algorithm. Pool takes a pinch of Salmon, a dash of Belich and a mix of positions including a spray on Moon and a thumbs up to Biggs to provide a picture against which or within which data might be considered. The cold reality of data collection meets the soft, feel-good sentiment of a slightly dislocated nationalism, a set of intellectual conceits, an emotional fall back position, a strange nationalism.

Several commentators imply that what the Indigenous group want, their ambitions in terms of their current circumstances, is the appropriate measuring stick for analysis.

These intellectual circumstances, this comfort constellation, underpin ethnicity as they do nationalism. It is correct to talk about the epistemology of sovereignty, as some commentators do in the Kukutai and Taylor collection. To be forever put up against foreign measuring sticks outside one's culture is distressing. Such might be seen more clearly in cases of the most dispossessed and this is one of the many values in this collection of papers as they add up to a kind of classic, if extreme, statement of estranged knowledge arrangements, of grievous and sometimes absurd knowledge conditions.

The critical link might now be between the Indigenous group and Big Tech. This might give us in the Maori context Ngai Tahu and Amazon, Rangitane and Alibaba or any kind of combination of iwi or pan iwi groups and Big Tech operators. Such is not considered by the commentators in the collection at least not in any depth. But with respect to data, especially data of buying and selling habits Big Tech has the data market, if such it be called, covered as it uses its algorithms to analyse that data. Iwi may well feel the need to be a critical part of such mixes of data and power.

Elsewhere (Cleave *Two Suns?* Five 2021) the present writer has argued that this is a matter of space in that if we consider, for example four types of space, neighbourhood space, national space, outer space and cyberspace we might see that Big Tech owns data in cyberspace and with the neighbourhood mesh Amazon owns data in the local space while Nation-states might own data in the space of the nation. The outer space situation is mixed with Big Tech now operating with its own data there. Somewhere in all of this are the ethnic groups, the minorities including Indigenous groups. A beginning point for analysis might be to consider where Indigenous groups stand in these four spaces.

The problem might be that the spatial template along with political, cultural and other templates commentators in the Kukutai and Taylor collection work off is that of the Nation-state which may or may not be treating the Indigenous group fairly. But that template, it is argued here and elsewhere, is suspect. We might come back to Ngai Tahu or another iwi comparing its data with that held by Amazon rather than or, at least, alongside the data held on that iwi by the New Zealand government or any other government for that matter. Amazon might hold as much or more data than the New Zealand Government but it might be data more attuned to the market world than the New Zealand

citizen lives in. And so the data picture of all groups including the Indigenous population in New Zealand held by the New Zealand state is one version while the data held of members of these group in Big Tech data banks might be another picture altogether. It's a question of assumed authority, of, if you like, colonialism and the mentality involved, of dependency, the nation state is assumed to have a better, a truer pictur

We are looking at system of hunting and gathering data. A search engine hunts for data which is gathered and hunted again as more data is searched and found. Crunched, and then farmed. This system works in terms of machine learning which dates, as mentioned above from Bayes. As mentioned above, the present writer has described this in a paper subtitled *The Algorithmic State, History and Theory* (Cleave 2021). Not addressed there but considered here is the blunt question, who is in charge?

Here is where it might get tricky for several kinds of scholars including those interested in ethnic affairs. One thing not to be assumed is that the Nation-states, especially the smaller nation states like, say, New Zealand, have effective control of data. To look at the situation of New Zealand Maori and data and to posit a strict control of data relating to Maori by the New Zealand state is not helpful when data to do with purchases, key movements on computers and, in fact, a multiplicity of data collection

processes is outside the purview of the state and effectively in the hands of Big Tech.

Then there is another matter to do with the last phrase, ‘the hands of big tech’ as this phrase assumes a personality behind it all. We have moved beyond state authority over data and then to a point beyond personal power. We may have talked about Bezos aka Amazon for a while, we may continue to do this even when Bezos removes himself progressively from day to day operations at Amazon as he is doing. But eventually we might have to consider that Amazon is a system removed from a charismatic personality, the most effective system we have perhaps of hunting, gathering and farming data. We thought there was a super brain a charismatic figure at the top but it was always to do with the algorithm known and developed in machine learning.

We might then proceed from these impersonal systems of data collection, analysis and use to ideas of political economy. Such ideas might relate to the arguments in the collection of writings about Indigenous data sovereignty by Kukutai and Taylor. However, matters of state and ethnicity are lost in the hunting, gathering and farming mix where the objective is to build better and better learning systems as the machine learning is set to do. The search engines do not have ethnic or national flags as those things have been known. The data is gathered and put together to create a world with its

own characteristics, its own map, and its own political economy. In that world it can be seen what sells where, who buys what, how often, in what quantity and a set of data aggregates, of data identities is created. This is more than a doppelganger, a version of the real world, it is its own thing. Forms, shapes of state and ethnicity are not, or at least not necessarily found in the way that they once were.

The use of data in the context of machine learning begs questions about identity and authority. We come back again to the question, ‘who is in charge?’ The key point of view, if there is one, is unclear. Data is sorted, links are made, data is sought related to those links and a learning model is set up relating to individual and group. Again it is a matter of hunting, gathering and farming data. But who or what runs the hunt, the gathering process and the farm?

There are some serious aspects to all this. The pain in stripping or absencing data about themselves from people is immense and in some respects the Indigenous data situation is indicative of the wider population now in their relation to Big Tech, a shape without a face that owns the data about people without their say. Regarding this immense pain which is reflected in each part of the Kukutai and Taylor collection as it is in the work of O’Neill and others regarding the wider population, we could talk about a psychic epidemic, of mental illness in the

form of mass psychosis. We could ask how we might step out of our minds to assess this situation. We could see a population of ourselves as weak and vulnerable persons in state of panic and lost in a childlike trust in the Nation state. We could consider all this as a pathological way of seeing reality.

We could talk about about powerless masses and a god-like few of powerful individuals who front the technology. We could talk of the addictive aspects of that technology. Of menticide. Over time we could talk of inter-generational trauma or of totalitarian psychosis which we had formerly reserved for fascism.

Or, to put it straight, can we handle data? If so, how?

We could come directly to questions of management. Is New Zealand of sufficient scale and capacity when it comes to data that it can effectively meet the very real and extremely important needs of the Maori? How many nation states have the scale and capacity to represent the data of Indigenous and minority groups let alone majority groups?

We are now looking at data and behaviour in a way not hitherto possible. Data is not simply a matter of record it is a resource for prediction. So we have data on what people buy and what they **might** buy, how that might be suggested. Going forward from that requires statistics from Nation-states like births and deaths to supplement data on clicks and searches neighbourhood meshes and other data

that has been hunted, gathered and farmed by Big Tech. However the latter might provide the bulk of the data involved.

We might talk, as the present writer does, about *The Data of Nations* (Cleave, Four 2020). And the phrase, ‘the data of nations’, might have several meanings. One could be that nations own data. Another implication might be that data transcends nations and redefines them. The phrase could refer to Adam Smith’s expression, ‘The Wealth of Nations’ or it could be implied that the data of nations is yet another resource of Big Tech. We might talk about a chain effect where there is the data of nations, the data held by Big Tech and the data of indigenous groups Whether nations or Big Tech are at the head of the chain might be debated. Kukutai and Taylor start their collection with the correct assertion that Indigenous groups did once and should again be at the head of the chain regarding their own data. Looking forward, what happens to data, how it is used in the context of machine learning and algorithms varies through the chain with Big Tech having superiority in such a context for the most part. Throughout such a chain we might think about the expression, ‘data is power’.

We are looking at data and the construction of an image. We could go back to Maggie Walker’s five ‘D’s above, of indigenous Australian data- disparity, deprivation, disadvantage dysfunction and difference in the Kukutai and Taylor collection. And

to her ideas of a ‘recognition space’, of ‘deficit data-problematic people’ and statistically informed pejorative stereotypes. This is like the data doppelganger thesis discussed above and throughout the *Two Suns?* series, especially in *Seven* (Cleave 2021) and coined by Sara Marie Watson (2014 *ibid*). People are made out to be something that they think they are not. Whereas the original sense of the term, ‘data doppelganger’, was meant to be regarding an individual, through Kukutai and Taylor’s collection there is the sense of distrust of the data deficit model by an ethnic group. Regarding the construction of an image, of an identity we might consider the data that Big Tech does and does not gather. For example Big Tech, at least in the form of Google will note a switch from English to Te Reo Maori. But unless explicitly instructed to do so there might not be an identification with waka, iwi hapu or whanau. The latter would need to be repeated in some way every time a person uses the te reo app. So we may have a retribalisation around data entry. Whakapapa is not noted unless explicitly stated so anyone New Zealand Maori, Cook Islands Maori or other speaker of a mutually intelligible Eastern Polynesian language, Pakeha or Other using Te Reo Maori immediately gets placed into a cluster, a community of interest based solely or primarily on their use of Te Reo.

At what point does a cluster of interest become a cluster of identity? Is this use of data a prompt for new tribal shapes? We might go back to the emergence of the King Movement in the nineteenth century or the Ratana Movement in the twentieth for comparable if not identical examples. The assembly of iwi in the Kingitanga allowed a new shape of identity as did the collection of morehu or remnants in the case of the Ratana movement. When the principle of data collection comes to centre on the speaking of Te Reo, when the algorithm runs to speakers before descendants, another shape emerges.

The field of data has changed and so has the farming of data. Where data was previously an historical item, with the algorithmic farming of data it became bits in a picture going forward. Before it was history and after it became prediction, there has been a move from the construction of past identity, say at the last census, to a steering of the future, say as to what you might buy next or where you might travel. You click on something in your computer, the search engines find something related and the next thing you know it's on your screen as a prompt. Click the prompt and there is now a construction of your interests to be followed. Your computer has a picture, a very rough sketch of you and has placed you in a community. That sketch of self and community is then embellished and directed in a

play of data involving search engines, algorithms and machine learning.

You are dragged into the now as the algorithms of social media operate. Data is not away in a vault or a library, data is at play on a screen that attracts the eye and mind with an idea of the user of the screen, the user's immediate needs and the apparently necessary and vital steps to take along a path now and in the future but, most of all, now.

There is a mix of data and adrenalin.

To return to the idea of Dsuz, data suzerainty discussed by Ian Pool in the Kukutai and Taylor collection we might think of the parastatal as comparable or in some ways identical to what is happening with Big Tech. There might be a comparison between, say, The East India Company in its day and Amazon today. There were technological advantages in the long day of the East India Company but the comparison fails when we consider such as similar to the technology, if so it be called, of machine learning. The East India Company might have had the best ships but ships do not think on their own, they do not hunt, gather, and farm data.

The East India Company kept records and accounts. They had data which they used effectively in their historical context and there were systems involved. But there was not the machine learning, the use of algorithms and other capacities of Big Tech to predict and inform the future. The shift is from

systems of record to systems of learning and the construction of plans and dreams for action.

This is not to take anything away from the sense in Ian Pool's last sentence in his article in the Kukutai and Taylor collection: "It would be tragic if this metamorphosed instead into neo-D suz under transnational corporate rule beyond the control of indigenous peoples or the polity in which they live", but we might proceed to ask in a serious way, what is the polity in which people live and how is that polity defined by data or by place or by system? Is the polity increasingly defined by a machine learning system fed by the data of individuals? Do we now have to face a situation where data feeds a learning machine which conjures a world for us to buy in, a world that gives us criteria to live by, a world to forget or displace the old tribes and create new ones?

Bibliography

- Bayes, Mr Price 1763 'An Essay towards Solving a Problem in the Doctrine of Chances' by the Late Rev Mr Bayes FRS Communicated by Mr Price *Philosophical Transactions of the Royal Society of London* 53: 370-418
- Bourdieu P 1984 *Distinction: a social critique of the judgment of taste*, Routledge, London.
- Walter and Andersen 2013
- Cleave, P 1989 *The Sovereignty Game. Power, Knowledge and Reading the Treaty* Victoria University Press
- Cleave P 2020 One -Four, 2021 Five-Eight *Two Suns? The Algorithmic State One-Eight* tekaharoa.com
- Cleave, P 2020 *Two Suns? Four The Data of Nations* tekaharoa.com
- <https://doi.org/10.24135/tekaharoa.v15i1.320>
- Cleave, P 2021 Two Suns? Seven, Data Doppelgangers and the Construction of the Digital Self, tekaharoa.com
- <https://doi.org/10.24135/tekaharoa.v17i1.352>
- Davis K E , Kingsbury B and Merry S E 2012 Indicators as a technology of global governance *Law and Society Review* 46 (1): 71-104 cited in Morhy
- Levi-Strauss. Claude 1966 *The Savage Mind* University of Chicago Press
- O'Neill, Cathy 2018 *The truth about algorithms* <https://www.youtube.com/watch?v=heQzqX35c9A>

- Rowley KG, O’Dea K, Anderson I, McDermott R, Saraswati K, Tilmouth R & Brown A (2008). Lower than expected morbidity and mortality for an Australian Aboriginal population: 10-year follow-up in a decentralised community. *Medical Journal of Australia* 188(5):283–7.
- Stewart-Harawira M (2005) *The new imperial order: indigenous responses to globalisation*, Zed Books, London
- Tournier C 2002 Don’t be on the other side of the digital divide: an overview of current health information initiatives impacting aboriginal peoples, Live address as part of the Aboriginal Health Information Symposium, Ottawa (11-13 February)
- Walter M and Andersen C 2013 *Indigenous Statistics and a quantitative research methodology*, Left Coast Press Walnut Creek CA.
- Watson Sara M. 2014 *Data Doppelgängers and the Uncanny Valley of Personalization* Why customized ads are so creepy, even when they miss their target *The Atlantic* JUNE 16, 2014