Appendix A: Canadian Social Science Indigenous Research (CSSIR) database data collection, coding protocol, and codebook

1. Research Questions

1. What are the primary methodological approaches used across social science disciplines to study Indigenous issues in Canada?

2. To what extent do Indigenous Canadian individuals or communities actively participate in social science research in their communities according to methodological approach or discipline?

2. Search strategy

The CSSIR search strategy included all articles published between 2005 and 2015 in peer-reviewed and grey literature based on a targeted search of Web of Science (core collection); EBSCOhost (Bibliography of Native North Americans and America: History and Life databases); ProQuest (Doctoral Dissertations and the Canadian Research Index); and Worldcat, which includes books. Collectively, these databases index journal articles, PhD theses, and books publishing social science research related to Indigenous peoples.

The search included filters for social science disciplines, including anthropology, development planning, economics, interdisciplinary social science, political science, social work and sociology. Studies from related disciplines, such as law, education, linguistics, and geography, were included if they explicitly focused on social conditions of Indigenous peoples (e.g. studies from human geography, sociology of education, comparative education, sociolinguistics, and social aspects and effects of law). We excluded psychology and health because research in those disciplines is normally funded by the CIHR. An exception are studies that explain health policy, which were included in the search.

What are key search terms?

Using Boolean operators, two general strategies were used to identify the sources to be included in the study. These strategies were tailored according to the particular characteristics of each database (WoS, EBSCO, ProQuest, etc.):

- Synonyms of Indigenous Peoples using truncated words, if applicable, and separated by OR, combined with (AND) words associated with Canada, and combined with (AND) research areas or disciplines. Exclusion of non relevant disciplines. Example:

  (("first nation*" OR indig* OR aborig* OR indian) AND (Canad* OR [all provinces' names]) AND ("social science*" OR [disciplines])) NOT (health OR "mental health" OR psych* OR "legal studies")

- Synonyms of Indigenous Peoples using truncated words, if applicable, and separated by OR, combined with (AND) all names of First Nations and Indigenous groups in Canada (separated by OR), and combined with (AND) research areas or disciplines. Exclusion of health and psychology. Example:

(("first nation*" OR indig* OR aborig* OR indian) AND (Inuit OR Métis OR [all names of First Nations in Canada]) AND ("social science*" OR [disciplines])) NOT (health OR "mental health" OR psych* "legal studies").

1. WoS search strategy: The search includes 3 steps (see Appendix 1 for syntax). All searches were conducted only in the relevant databases within the WoS core collection, including the Social Sciences Citation Index and Conference Proceedings Citation Index – Social Sciences and Humanities:

    -   Combination of search by topic (TS) using synonyms of Indigenous Peoples (indigen* OR aborigin* OR "first nation*" OR indian) AND (Canada OR all provinces' names). Combination of the previous search topic with a search by research area (SU) that includes all areas relevant for this project. Restriction of the search by language (English), document type (article, book, book chapter or proceeding paper), and publication date (2005-2015).

    -   Combination of search by topic (TS) using synonyms of Indigenous Peoples (indigen* OR aborigin* OR "first nation*" OR indian) AND names of First Nations and Aboriginal Peoples . Combination of the previous search topic with a search by research area (SU) that includes all areas relevant for this project. Restriction of the search by language (English), document type (article, book, book chapter or proceeding paper), and publication date (2005-2015).

    -   Combination of steps 1 OR 2 excluding all irrelevant WoS categories and research areas (see Appendix 1).

Searches in WoS produced 596 results

2. EBSCO search strategy: The team compiled a list of thirty-two journals relevant for this project. Based on the databases in which these journals were indexed, two databases within EBSCO were selected to conduct the search (America: History and Life and Bibliography of Native North American). The following search was conducted in each of these two databases:

-   America: History and Life (see Appendix 1 for syntax).

    -   Combination of search by subject (SU) using synonyms of Indigenous Peoples (indigen* OR aborigin* OR "first nation*" OR indian) AND (Canada OR all provinces' names). Exclusion of irrelevant subjects (NOT SU) (see Appendix 1). Restriction of the search by publication date (2005-2015), document type (Article, Book, Book Chapter, Conference Paper, Dissertation, Report), and language (English). Narrowing search to relevant journals (see Appendix 1).

    -   Combination of search by subject (SU) using synonyms of Indigenous Peoples (indigen* OR aborigin* OR "first nation*" OR indian) AND names of First Nations and Aboriginal Peoples. Exclusion of irrelevant subjects (NOT SU) (see Appendix 1). Restriction of the search by publication date (2005-2015), document type (Article, Book, Book Chapter, Conference Paper, Dissertation, Report), and language (English). Narrowing search to relevant journals (see Appendix 1).

    -   Combination of steps 1 OR 2.

Since health was one of the subjects excluded from the previous searches, another search was added to capture health policy articles:

- Combination of search by subject (SU) using synonyms of Indigenous Peoples (indigen* OR aborigin* OR "first nation*" OR indian) AND (Canada OR all provinces' names) AND subject (SU) health AND subject (SU) policy. Restriction of the search by publication date (2005-2015), document type (Article, Book, Book Chapter, Conference Paper, Dissertation, Report), and language (English).

- Combination of search by subject (SU) using synonyms of Indigenous Peoples (indigen* OR aborigin* OR "first nation*" OR indian) AND names of First Nations and Aboriginal Peoples AND subject (SU) health AND subject (SU) policy. Restriction of the search by publication date (2005-2015), document type (Article, Book, Book Chapter, Conference Paper, Dissertation, Report), and language (English).

- Combination of steps 1 OR 2.

- Bibliography of Native North Americans (see Appendix 1 for syntax)

  - Combination of search by subject (SU) using synonyms of Indigenous Peoples (indigen* OR aborigin* OR "first nation*" OR indian) AND (Canada OR all provinces' names). Exclusion of irrelevant subjects (NOT SU) (see Appendix 1). Restriction of the search by publication date (2005-2015), document type (academic journal, book), and language (English). Narrowing search to relevant journals (see Appendix 1).

  - Combination of search by subject (SU) using synonyms of Indigenous Peoples (indigen* OR aborigin* OR "first nation*" OR indian) AND names of First Nations and Aboriginal Peoples. Exclusion of irrelevant subjects (NOT SU) (see Appendix 1). Restriction of the search by publication date (2005-2015), document type (academic journal, book), and language (English). Narrowing search to relevant journals (see Appendix 1).

  - Combination of steps 1 OR 2

  - To capture health policy articles, a search was added similar to the one conducted with America: History and Life

Searches in Ebsco produced 521 results

3. ProQuest Search strategy: Two databases were searched in Ebsco: Canadian Research Index and ProQuest Dissertations & Thesis A&I

- Canadian Research Index

  - Combination of search by subject (SU) using synonyms of Indigenous Peoples (indigen* OR aborigin* OR "first nation*" OR indian) AND (Canada OR all provinces' names) AND subject synonyms of research (research OR survey OR "case stud*" OR stud*). Exclusion of anual reports, progress reports, performance reports, and plans. Restriction of the search by language (English) and publication date (2005-2015). Exclusion of non relevant subjects.

  - Combination of search by subject (SU) using synonyms of Indigenous Peoples (indigen* OR aborigin* OR "first nation*" OR indian) AND names of First Nations and Aboriginal

Peoples AND subject synonyms of research (research OR survey OR "case stud*" OR stud*). Exclusion of anual reports, progress reports, performance reports, and plans. Restriction of the search by language (English) and publication date (2005-2015). Exclusion of non relevant subjects.

- Combination of steps 1 OR 2

- To capture health policy articles, a search was added similar to the one conducted with EBSCO.

- ProQuest Dissertations & Thesis A&I

  - Combination of search by subject (SU) using synonyms of Indigenous Peoples (indigen* OR aborigin* OR "first nation*" OR indian) AND (Canada OR all provinces' names) AND subject social sciences. Exclusion of irrelevant subjects. Restriction to doctoral dissertations, language (English), and publication date (2005-2015).

  - Combination of search by subject (SU) using synonyms of Indigenous Peoples (indigen* OR aborigin* OR "first nation*" OR indian) AND names of First Nations and Aboriginal Peoples AND subject social sciences. Exclusion of irrelevant subjects. Restriction to doctoral dissertations, language (English), and publication date (2005-2015).

  - Combination or steps 1 OR 2

  - To capture health policy articles, a search was added similar to the one conducted with EBSCO.

Searches in ProQuest produced 139 results.

4. Worldcat: Search with names of First Nations and Aboriginal Peoples were not feasible in Worldcat. The browser did not accept so many search terms.

- Combination of search by subject (SU) using synonyms of Indigenous Peoples (indigen* OR aborigin* OR "first nation*" OR indian) AND (Canada OR all provinces' names) AND relevant subjects ; exclusion (NOT) of irrelevant subjects . Limited by publication date (2005-2015), language (English), number of libraries that held the item (50 or more), document type (books), audience (not juvenile), content (not fiction). Restriction to university libraries with graduate programs (Master or PhD) in Indigenous Studies

- To capture health policy books, a search was added similar to the one conducted with EBSCO

Search in Worlcat produced 112 items

5. Unindexed journals: From the list built by the team, we handsearched articles in non-indexed journals.

This search produced 65 results.

6. Grey literature: in order to complement the search for grey literature, we handsearched in the publication section of different NGOs and Indigenous organizations (see Appendix IV). The list of organizations was compiled by the team and through the iPortal (Indigenous Studies Portal Research Tool) browser that allows the identification of organizations that conduct

Indigenous studies. iPortal was also handsearched for reports and documents based on research using the following combination:

Description=(research OR survey OR stud*) AND Resource Type=(chapters OR documents & presentations OR ebooks).

This search produced 70 results.

3. Selection criteria and strategy

- Must have <u>Canadian</u> Indigenous/Aboriginal focus

- Included fields: anthropology, business, development planning, economics, education, geography, interdisciplinary social science, library and information science, linguistics, political science, social work and sociology (additional interdisciplinary vary among databases. Some examples are women studies, area studies, youth studies, etc.)

- Excluded fields: psychology, mental health, health (excepting studies that explain or interpret health policy or politics), legal and law development studies, history, educational studies (curriculum and pedagogy), archaeology.

To select the items that would be included in the CSSIR, two reviewers checked the title and abstract of each one of the 1,503 items collected in the search phase. These reviewers assessed each item with yes, maybe, or no. In case of disagreement in the assessment, a third reviewer resolved the conflict. With this strategy, a total of 801 items were selected and the remaining and 702 were rejected.

4. Study coding

The 740 unique items identified by the search were coded automatically for the following indicators (variable names in parenthesis) using the information from the source databases (WoS, Ebsco, ProQuest, Worlcat, etc.):

- **Author and Journal keyword:** Where available, the database retains all author keywords (if provided as a separate field) as well as journal or indexing service keywords in two separate variables.

- **Number of authors (numaut)**

- **Author's (probable) gender (A1_female – A5_female)**: We used genderizeR (Wais 2016) and genderize.io (https://genderize.io/) to estimate the probability that an author was female. Genderize.io uses social media and other administrative datasets that combine first names with user genders to predict whether a first name is likely to be associated with someone who identifies as male or female. Based on these predicted probabilities, probabilities greater than 0.8 were coded as female and less than 0.2 were coded male. In instances where the automated coding did not return a predicted probability or the probability was between 0.2 and 0.8, no gender was coded for an author. This method has been used in other recent studies (Teele & Thelen 2017; Sumner forthcoming), including one that estimated the error rate of genderize.io incorrectly assigning a gender based on first name to be about 2% when compared to manual coding of gender by research assistants. We acknowledge that both methods (to varying extents) are based on the problematic assumption that gender is a binary construct.

- **Open access:** A student assistant used the Directory of Open Access Journals (doaj.org) and targeted web searches to determine whether journal titles are open access. This does not account for open access articles published non-open access journals.

Student research assistants coded 501 articles for various characteristics. Books, theses, some journal articles, and other monographs (e.g., grey literature) were not coded due to inconsistent availability. The student assistants were trained to code the characteristics below by separately coding 5-10 articles. Members of the team then met to discuss differences in coding and to clarify concepts and methods. Once the research assistants were trained, due to limited resources and a desire to code as many items as possible, each article was coded by one assistant. However, assistants also indicated whether the article was easy or difficult to code. Those marked difficult, were then coded by another assistant and differences were resolved by a third member of the team. The key team members met weekly during the coding period to discuss any concerns or issues that arose during the coding process.

- **Affiliation of each author (A1_orgtype to A5_orgtype)**: The type of organization each of the first 5 article authors are affiliated with: educational institutions, colonial government, Indigenous government, NGO, unknown.

- **Author's self-identified Indigenous status (A1_sid to A5_sid)**: The self-identification of the first 5 article authors, if known: Indigenous, non-Indigenous, unknown.

- **Place name (place):** Proper names of study location, when available in the text (e.g., province, community, treaty territory, town, Nation, etc.)

- **Location type**:
    - **(Urban)**: yes/no
    - **(On-reserve)**: yes/no

- **Comparative (compare)**: comparative with group outside Canada (yes/no)

The methodological approach categories have **four** dimensions, each one with multiple yes/no variables as follows:
1. **Use of Indigenous language during the research (ind-lang):** The study uses Indigenous language(s) to conduct any part of the research, including research proposal, data collection, analysis or dissemination products. Yes or No.
2. **Epistemology:**
    - **Positivist/post-positivist (epi-pos)**: Assumes reality can be measured, although the measurement may be imperfect.
    - **Constructivist (epi-const)**: Assumes that reality is socially constructed.
    - **Critical (epi-critical)**: Individual and community reality is constructed by historical or institutional oppression.
    - **Constructionist/post-modern (epi-pm)**: Assumes reality is individually, internally constructed by experience.
    - **Indigenous (epi-ind)**: Assumes reality is relational and constructed through relationships between individuals, nature, etc.
3. **Data Sources:**
    - **Experiments (data-experiment)**: Participants complete tasks or answer questions that enable data collection. Can be in a lab, in the field, or on a survey.

- **Surveys (data-survey)**: A series of questions, often with a closed list of response options, administered in person, online, by mail, or over the telephone. Synonymous with questionnaire.
- **One-on-one dialogue (data-1_1)**: One-on-one conversations with participants, often including open-ended questions posed by the researcher, including interviews. Also included are "postcolonial Indigenous interview methods" differ from social science interview methods because the researcher does not guide the interview but facilitate the construction of the interviewee's story, and there is an explicit effort to neutralize power imbalances. All data sources in this category are based on some form of one-on-one interaction between the researcher and an informant.
- **Experience (data-experience)**: Participating in or attending events, meetings, or other activities and recording observations of actions, etc. Common method of data collection in ethnographic studies, and may include use of audio/video recordings. May be called participant observation, reflexive autoethnography, or ceremonies.
- **Group dialogue (data-group)**: Group interviews or conversations, often including open-ended questions and discussion. Would include Indigenous forms of collective knowledge sharing with a researcher, sometimes called a sharing circle.
- **Primary sources (data-psource)**: Archival documents/media or collections, government-generated data/statistics, administrative datasets from government, "big data", newspapers, photos, ethnographic films, diaries, audio recordings, or transcriptions that provide the raw material for a researcher to answer their question. Usually these sources were not created for the purpose of the research. Can include online materials.
- **Secondary sources (data-ssource)**: Grey literature (government, NGO reports), scholarly literature, etc. that engage in analysis of an issue or subject.
- **Reflexive sources (data-rsources)**: Directed by or in consultation with researchers, participants record their experiences over time in text or other media (e.g., video or photography). These sources are created specifically for the research. Participatory GIS and counter-mapping would be a form of reflection or capturing of geolocation data or socially constructed maps of spaces. Other types of reflexive sources include art-based methods, and symbol-based reflections.
- **Other (data-other)**: Anything else that doesn't fit one of the categories above.

4. **Methods of data analysis:**
   - **Quantitative descriptive (meth-qdesc)**: Use of descriptive statistics (means, frequencies, cross-tabulations) to <u>describe a sample or population</u>. Tables or figures will often be simple and <u>easy for non-specialist to understand</u>. Some types of <u>network analysis</u> are descriptive and aim only to describe the network (not attribute cause/effect). May mention particular tools, such as Excel, SPSS, Stata, or R. This could include Indigenous Statistics, or descriptive statistics from an Indigenous perspective or worldview.
   - **Quantitative inferential (meth-qinf)**: Use of <u>inferential statistics</u> to establish correlations or causation, usually including one dependent variable and several independent variables (i.e., multivariate) and usually some form of <u>regression-based</u> method. Results will include mention of <u>statistical significance</u> and include "models" in tables or figures of results. Surveys, experiments, and administrative data are often used in these analyses. Some types of <u>network analysis</u> aim to understand relationships between network structure or position and particular outcomes or dependent variables.

Quasi-experimental methods including "matching" methods would fall into this category, too. May reference particular software tools, such as: Excel; MLwiN; Mplus; Python; R; SAS; SPSS; Stata; Winbugs.

- **Interpretive ethnography (meth-inteth)**: Immersion of the researcher in the participants' environment to identify and understand their culture, identity, goals, etc. Key is to uncover meanings. Includes interpretive, reflexive, auto, and critical versions. May mention tool such as: ATLAS.ti; Maxqda; NVivo; Transana; Computer Aided Qualitative Analysis Software (CAQDAS).
- **Positivist ethnography (meth-poseth)**: Immersion of the researcher in the participants' environment to test hypotheses or understand causal processes. Aims to be objective and descriptive. Not as interested in the construction of meaning or culture but instead in explaining why particular outcomes occur. May mention tools such as: ATLAS.ti; Maxqda; NVivo; Transana; Computer Aided Qualitative Analysis Software (CAQDAS).
- **Qualitative content analysis (meth-cont)**: methods that focus on analysing the functional and sense-making properties of written and oral language, visual imagery, or other forms of cultural media. They assess the internal and external consistency of the speaker and message. These methods include textual, visual, discursive, conversational, thematic, framing, framework, and narrative analysis. May mention tools such as: ATLAS.ti; Maxqda; NVivo; Transana; Computer Aided Qualitative Analysis Software (CAQDAS).
- **Discourse analysis (meth-disc):** analysis of how meanings are constructed by texts, institutions, communities, etc. as a means to understand underlying assumptions/biases/power dynamics. In this context, most often refers to critical or post-modern approaches (i.e., post-colonial, constructionism/constructionist, Foucauldian analysis). Will often not include primary sources of evidence and instead operate at a high level of abstraction or observation.
- **Qualitative comparative case analysis (meth-case)**: case-based methods that aim at tracing processes that lead to specific outcomes by identifying necessary and sufficient causal conditions. Can include comparisons over time, or across different time periods, or historical analyses. Though cases are considered "whole," they are also often disaggregated into independent and dependent variables. Boolean and QCA are two variations on comparative case studies that use Boolean algebra and logic to deduce patterns of cause and effect.
- **Inductive analysis (meth-induc)**: methods that analyze empirical qualitative data to develop theory inductively without relying on preconceived theories. May be called grounded theory. Often combined with interpretive ethnography or discourse analysis. May mention tool such as: ATLAS.ti; Maxqda; NVivo; Transana; Computer Aided Qualitative Analysis Software (CAQDAS). Will differ from discourse analysis in that it will usually include a corpus of documents, experience, fieldwork or other observations that are analyzed inductively to develop a theoretical (descriptive, interpretive or even causal) analysis.
- **Participatory methods of analysis (meth-partic)**: This may include participant action research or other forms of community-engaged or community-based participatory research, which include active participation of the community. The researcher serves as a facilitator or catalyzer of the research process. It can also include other participatory methods whereby "participatory research approaches enable the colonized… to

collectively share and analyze their knowledge, life experiences, and conditions and to use Indigenous knowledge as a frame of reference to plan and act" (Chilisa 2012, 225). The common characteristics of these methods are the central role that the subject of research takes in not only creating source material (e.g., video diaries, sharing circles) but also an active role in the analysis and interpretation of the material. May also include performances and other atypical research creation objects (see SSRHC on "research creation"), some of which may not appear in article/book/document form (and therefore will be systematically underreported in our study).

- **Agent-based modelling/simulations (meth-mod)**: Methods that create simulations to mimic hypothetical human behaviour to understand how context or interactions lead to different outcomes. Often used in biological sciences, epidemiology, or ecology to model interactive behaviours.

- **Other (meth-other)**: Anything not described above

Indigenous participation categories describe a continuum of level of involvement that starts with no participation at all and ends with Indigenous-led research:

1. Participation (ICpartic): coded with one of the following values

- **No human subjects involved (none):** Theoretical and/or conceptual studies or research based on secondary sources (e.g. existing surveys, document analysis, etc.). Studies in this category do not establish any relationship with Indigenous communities. Outcomes are mainly peer-reviewed publications and participation in scholarly conferences. Sometimes, diffusion channels can also include publications in op-eds, magazines, specialty industry journals (non peer-reviewed), technical reports, blogs, websites, podcasts, on social media, and participation in media interviews.

- **Omitted participation (omitted):** Studies that involve human subjects but omit the involvement of Indigenous individuals and the review and approval from Indigenous communities. Diffusion channels are similar to those used by studies with no human subjects involved.

- **Acknowledge (acknowledge):** Studies in this category limit the participation of Indigenous communities to the approval of ethics protocol. Indigenous communities do not engage or participate actively in the project but simply acknowledge it and register no objection to it. Diffusion channels are similar to the previous two categories.

- **Engage (engage):** Minimal involvement of Indigenous communities in the project focused primarily on sharing information. This involvement can include having Indigenous Peoples mainly as informants and/or obtaining some feedback on research design, research questions, choice of methodology, analysis, and/or decisions regarding dissemination. The relationship with Indigenous partners is limited to keeping them informed, listening to and acknowledging the knowledge, goals and concerns of the community. Communication can be occasional or regular, but no clear governance structure is present. Diffusion channels are similar to the previous categories but partners are more intimately involved in alternative outputs, such as workshops, roundtables, community meetings or presentations, and content verification with partners

- **Collaborate (collaborate):** Strong, entrenched involvement. Partner with Indigenous participants in each aspect of the research process, including the development of

research questions, research design, and analysis of findings. Although studies are led by non-Indigenous organizations, there is a strong relationship with partners, which often includes participation in governance structures or research advisory boards. Partner point of view is clearly incorporated into the research. Publications in scholarly and lay outlets come out of this research, but the community-centered outputs are much more of a focus, such as, community reports and presentations, co-authorship and presentations with community members, results to partners, and member checking.

- **Empower (empower):** Very strong, entrenched involvement. Empower participants or Indigenous research partners to leverage the knowledge to have an impact in their environment and continue to carry on research that benefits them by imparting new skills, knowledge and/or resources. Studies are initiated by Indigenous organizations and/or derive from a pre-existing partnership between Indigenous and non-Indigenous communities with fully shared leadership. Indigenous partners actively sit on governance and advisory boards. Indigenous partner perspectives are fully reflected in the research and its outputs and the entire research process is accountable to the Indigenous community. Diffusion channels are similar to those in the Collaborate category but they go beyond research outcomes to build capacity in the community. Sometimes this is done through training and learning of new skills, advocacy, knowledge transfer, and increased awareness.

2. Indigenous-led studies (ICled): (yes/no) Studies in which Indigenous peoples and communities initiate and have leadership of the research process, which means they have greater power than non-Indigenous researchers or organizations. These studies go beyond research conducted by individual Indigenous researchers in educational or non-educational institutions.