

Disaggregated Data in Canada: Leveraging Citizen Data to Complement Official Statistics

Tony Labillois

Regional Statistics Conference 2026, University of Malta

Abstract

In Canada, the growing emphasis on disaggregated data reflects a broader shift toward understanding the diverse and intersecting realities of populations and communities. Led by Statistics Canada, initiatives such as the Disaggregated Data Action Plan (DDAP) or the Quality of Life Indicators aim to expand the availability, accessibility, and use of detailed, intersectional data to support more equitable and evidence-based decision-making.

While official statistics remain foundational—offering methodological rigor, comparability, and public trust—they face persistent challenges related to timeliness, granularity, and the representation of certain populations. At the same time, a growing ecosystem of citizen-related data—including administrative records, digital service interactions, and community-generated data—creates new opportunities to complement official statistics and enrich the analytical picture of Canadian society and the economy.

This paper examines how these diverse data sources can be brought into dialogue. Drawing on some Canadian examples such as health and socio-economic data during COVID-19, housing and homelessness measurement, Quality of Life Indicators, and Indigenous data contexts, it highlights how citizen data can help address gaps, improve coverage of underrepresented groups, and enable more granular and intersectional insights. Particular attention is given to the role of data linkage, interoperability, and evolving statistical standards in enabling such complementarities.

The paper also reflects on key considerations that shape this evolving data landscape, including data quality, representativeness, privacy, consent, and trust, as well as broader governance and stewardship frameworks. These issues are examined in light of Canada's policy context, including human rights, equity, and inclusion objectives embedded in disaggregated data strategies.

Rather than prescribing solutions, the document aims to contribute to an ongoing discussion on how official statistics and citizen data can jointly support more

inclusive, timely, and policy-relevant insights in an increasingly complex data ecosystem.

Keywords: administrative data; citizen-data; data disaggregation; data governance; data integration; data interoperability; equity; human rights; inclusion; official statistics; record linkage

1. Introduction

The statistical picture of a country is necessarily a construction. It depends on what is measured, how concepts are defined, which populations are reached, which variables are collected, at what geographic level information is available, and how the resulting data are processed and communicated. For many decades, national statistical systems have provided the principal infrastructure through which governments, researchers, businesses and the public understand economic, social, demographic and environmental conditions. Official statistics remain indispensable to democratic societies because they provide impartial and methodologically grounded information to support public decision-making and informed public debate (United Nations, 2014).

Yet the statistical portrait produced through conventional systems can obscure important differences between population groups. A national average may be accurate while simultaneously concealing substantial disparities by gender, age, Indigenous identity, racialized identity, disability, income, geography or other characteristics. The challenge becomes even more significant when these characteristics intersect. The experience of a young Black woman, for example, cannot necessarily be inferred by separately examining statistics for young people, Black populations and women. Intersectionality therefore requires statistical systems to move beyond the simple comparison of single characteristics toward a more nuanced understanding of how multiple social and structural factors interact (Crenshaw, 1989; Statistics Canada, 2026a).

Canada has responded to this challenge through a substantial expansion of work on disaggregated data. Statistics Canada's Disaggregated Data Action Plan (DDAP) represents a whole-of-government effort to improve the availability, accessibility and use of detailed data on diverse populations. Its objectives include addressing systemic racism, gender gaps and other inequities, while strengthening the evidence available for policy and programme development. The DDAP initially focused on four employment-equity population groups—Indigenous Peoples,

women, racialized populations and persons with disabilities—while extending disaggregation, where relevant and feasible, to other populations and characteristics.

At the same time, the Canadian data environment is changing beyond the boundaries of the traditional national statistical system. Governments increasingly hold administrative records; people interact with digital public services; community organizations collect information about populations they serve; non-profit organizations conduct surveys; academic researchers develop new data collections; private organizations generate large volumes of transactional and digital information; and citizens themselves increasingly participate in the production of data about their communities.

This changing environment creates an opportunity to expand, rather than a risk to weaken, the role of official statistics. It suggests a broader question: **how can official statistics and other forms of data be brought into a productive relationship so that each contributes what the other may lack?**

This paper explores that question in the Canadian context. Its central argument is deliberately modest. Citizen-related data should not be regarded as a substitute for official statistics. Nor should every alternative data source be treated as suitable for statistical production. Rather, carefully governed complementary data ecosystems can help official statistics address specific gaps in timeliness, granularity, geographic coverage, population coverage and understanding of lived experience.

2. From averages to diversity: the case for disaggregated data

Disaggregation is fundamentally about making differences visible. Aggregate statistics answer questions about populations as a whole; disaggregated statistics help answer questions about how outcomes are distributed across populations and places.

This distinction has direct implications for equity. If a policy affects different groups differently, an aggregate indicator may fail to reveal the distributional consequences of the policy. Conversely, data that identify differences across groups can help policymakers determine where barriers exist, which populations are disproportionately affected and whether interventions are producing more equitable outcomes.

This logic is increasingly reflected in Canada's policy environment. The federal government's Gender-Based Analysis Plus (GBA Plus) framework recognizes that gender interacts with other characteristics and circumstances, including age, disability, education, ethnicity, race, geography, sex, religion, economic status and language. The use of disaggregated data is consequently not simply a statistical refinement; it is part of a broader effort to understand differential experiences and outcomes.

The human-rights dimension reinforces this argument. A human-rights-based approach to data emphasizes participation, non-discrimination, equality, transparency and accountability. The Office of the United Nations High Commissioner for Human Rights has specifically argued that disaggregated data are necessary to identify groups that are being excluded or experiencing discrimination and that data collection should involve rights-holders, including Indigenous Peoples and persons with disabilities (OHCHR, 2018).

Disaggregation is therefore relevant not only to statistical accuracy but also to the ability of governments and societies to recognize inequalities.

2.1 Intersectionality

Intersectionality adds a further layer. Developed initially as a framework for understanding how systems of discrimination interact, intersectionality challenges approaches that treat social characteristics as independent dimensions (Crenshaw, 1989).

For statistical systems, the practical implication is important. It is not sufficient to know that women experience a particular outcome differently from men, or that racialized populations experience a particular outcome differently from non-racialized populations. In some circumstances, the relevant policy question concerns the intersection of these and other characteristics.

Statistics Canada's DDAP explicitly incorporates this principle. Its four guiding principles require data to be disaggregated to the lowest feasible level of population detail, analysis to focus on intersectionality rather than only binary combinations, approved statistical standards to be used across programmes, and data to be released at the lowest feasible geographic level, subject to quality and confidentiality considerations.

Intersectionality consequently affects the entire statistical value chain: planning, conceptualization, data collection, processing, linkage, analysis and dissemination. It also changes the question of what constitutes a data gap. A dataset may provide

information on a population while still failing to provide meaningful information about important subgroups within that population.

3. Canada's institutional and statistical context

Canada's federal structure makes the question of complementary data particularly significant. Statistics Canada occupies a central position within the national statistical system, but statistical information is produced and held across federal departments, provincial and territorial statistical organizations, municipalities, Indigenous governments and organizations, academic institutions, non-profit organizations and private-sector entities.

This creates both opportunities and coordination challenges. The statistical system is no longer adequately represented by a single institutional producer. Instead, Canada increasingly operates within a **data ecosystem** in which multiple actors are producers, users, intermediaries and stewards of data.

Statistics Canada's own work increasingly reflects this reality. The agency has expanded its use of administrative data and data linkage, developed partnerships with other organizations, increased engagement with communities and introduced new approaches to data access and dissemination. The DDAP itself is explicitly a whole-of-government initiative, and its implementation involves engagement with departments, communities, researchers and other partners.

This broader ecosystem also includes organizations that traditionally would have been viewed primarily as data users. Non-profit organizations, academic institutions, community organizations and private companies increasingly produce information that can be analytically valuable. In a federal context, this plurality of data producers makes institutional alignment, coordination and access particularly important to the quality and timeliness of national statistics (Di Gennaro Splendore and al., 2026). Some of this information is collected intentionally for statistical or research purposes; other information emerges from the normal operation of programmes, services or platforms.

The distinction matters. The United Nations' emerging framework for citizen data defines citizen data more narrowly as data originating from initiatives in which citizens initiate or are sufficiently engaged in the design and/or collection stages of the data value chain (United Nations, 2025). This is different from administrative data, which may be generated as a by-product of government or organizational operations.

For the purposes of this paper, therefore, **citizen-related data** is used as a broad practical term encompassing administrative, digital, community-generated and citizen-generated sources. This avoids conflating fundamentally different forms of data while recognizing that they can all potentially contribute to a richer statistical ecosystem.

4. The Disaggregated Data Action Plan

The DDAP provides an important Canadian institutional foundation for this discussion. Established through Budget 2021, the initiative was funded initially with \$172 million over five years, with ongoing funding subsequently provided. Its purpose is to strengthen the collection, analysis, availability and publication of disaggregated data and to improve the representation of diverse populations in Canadian statistics.

The DDAP is organized around four guiding principles.

First, data and analysis should be disaggregated to the lowest possible level of population detail while respecting quality and confidentiality. Second, analysis should focus on intersectionality rather than simply examining pairs of characteristics. Third, Statistics Canada's approved standards should be used consistently. Fourth, data should be released at the lowest possible geographic level, subject to the necessary safeguards.

These principles are supported by five broad areas of activity: expanding disaggregated data assets; increasing intersectional and longitudinal insights; improving access; developing and promoting statistical standards; and enhancing engagement and communication.

The progress is substantial. The DDAP has contributed to new datasets and analytical products covering diverse populations, new statistical standards, enhanced access to data and increased collaboration with partners. In 2022–2023, 46% of Statistics Canada's analytical products included disaggregated information on specific population groups, and almost 200 research articles, infographics and visualization products were released under the programme's broader work.

The subsequent development of the programme has reinforced the cultural dimension of this work. Statistics Canada has expanded training, analytical capacity and engagement with federal, provincial, territorial, municipal, Indigenous, academic and community partners. Its 2024–2025 reporting emphasizes that the

DDAP has become part of a broader cultural shift toward routinely considering disaggregation and intersectionality in statistical work.

The significance of the DDAP for the present discussion is therefore broader than the production of additional tables. It represents an attempt to change the statistical system from an approach that primarily describes the population as an aggregate toward one that deliberately asks **whose experiences are visible, whose are missing, and at what level of detail the differences matter.**

5. The Quality of Life Framework: broadening what is measured

A parallel development is Canada's Quality of Life Framework. Developed through federal consultations and introduced through Budget 2021, the framework recognizes that economic indicators alone cannot provide a complete account of well-being.

The framework brings together indicators across prosperity, health, society, environment and good governance, with cross-cutting lenses of fairness and inclusion and sustainability and resilience. It also incorporates subjective measures such as life satisfaction, sense of meaning and purpose, and future outlook.

This represents an important conceptual development for official statistics. It shifts the statistical question from simply measuring economic activity to understanding how economic, social, environmental and institutional conditions affect people's lives.

The framework also creates a natural space for complementary data. Official indicators can describe measurable conditions—for example, employment, housing affordability, health outcomes or food insecurity. Citizen and community data can potentially provide additional information about perceptions, priorities, experiences and emerging needs.

The distinction is important. A national indicator may show the prevalence of a condition, while community-generated information may help explain how that condition is experienced locally. A statistical measure can tell us how many people are affected; participatory or citizen-generated data may help reveal what the affected population considers important about the experience.

The Quality of Life Framework has itself evolved toward greater granularity. Recent developments include more geographically disaggregated information and increased attention to demographic breakdowns. Statistics Canada's 2024 review of

the framework highlighted demand for regionally disaggregated data and detailed breakdowns by characteristics such as sex, gender, rural or urban status and Indigenous identity (Statistics Canada, 2025a).

This illustrates the central theme of this paper: **the statistical picture becomes richer when different forms of evidence are designed to answer complementary questions.**

6. The strengths and limitations of official statistics

The case for complementary data should not be confused with a critique of official statistics as an outdated model. Official statistics have distinctive strengths.

First, they are produced within institutional and methodological frameworks designed to ensure relevance, accuracy, timeliness, accessibility, coherence and interpretability. Statistics Canada has a long-established quality assurance framework that explicitly recognizes that these dimensions can sometimes be in tension (Statistics Canada, 2024b).

Second, official statistics provide standardized concepts, classifications and methodologies. This supports comparability across time, geography and population groups.

Third, official statistics operate within legislative and institutional safeguards concerning confidentiality, privacy and professional independence. These characteristics contribute to public trust and make official statistics a central component of democratic accountability (Di Gennaro Splendore, 2024; Labillois, 2026).

Nevertheless, official statistics face practical constraints. Traditional surveys can have long collection-to-publication cycles. Sample sizes may be insufficient for small populations or detailed geographic areas. Increasing disaggregation can create small-cell confidentiality and quality problems. More extensive survey collection can also impose respondent burden and financial costs.

Statistics Canada's recent work illustrates the methodological response to some of these limitations. Small-area estimation, for example, can combine Labour Force Survey information with Employment Insurance statistics and demographic projections to improve estimates for small communities where direct survey estimates are imprecise (Statistics Canada, 2025b).

The limitations are therefore not evidence that official statistics are inadequate. Rather, they show why statistical systems increasingly require multiple methods and multiple sources.

7. Citizen-related data as a complementary source

The concept of citizen data is gaining international recognition. The United Nations Collaborative on Citizen Data has developed the Copenhagen Framework, which was endorsed by the UN Statistical Commission in 2025. The framework defines citizen data around meaningful citizen engagement in the design and/or collection of data and emphasizes its potential to supplement national statistical systems, fill data gaps and strengthen the representation of underrepresented populations (United Nations, 2025).

Citizen-related data can take several forms.

Administrative data arise from the delivery of public programmes and services. They can provide large-scale, longitudinal and often highly detailed information.

Digital service interactions can generate information about how people access government or community services and where barriers occur.

Voluntary and community surveys can provide information on experiences, perceptions and priorities that may not be captured in conventional surveys.

Crowdsourced data and online platforms can provide rapid information, especially during periods of crisis or rapid social change.

Community-generated data can be particularly valuable when communities have knowledge of populations or circumstances that are difficult for conventional statistical programmes to reach.

The principal advantage is not necessarily that these sources are “better” than official statistics. Rather, they may be **different**. They can be faster, more local, more continuous or closer to lived experience.

That difference is precisely what creates the possibility of complementarity.

8. COVID-19: an illustration of complementarity under pressure

The COVID-19 pandemic demonstrated the importance of timely and flexible data systems. Conventional statistical collection was disrupted, while demand for information increased dramatically.

Statistics Canada responded by using alternative collection methods, including crowdsourcing. By September 2020, more than 500,000 Canadians had voluntarily participated in Statistics Canada's crowdsourcing activities during the pandemic. These included initiatives on the impacts of COVID-19, postsecondary students, mental health, personal safety, trust in institutions, and families and children (Statistics Canada, 2020b).

The crowdsourcing results provided rapid information about how Canadians were experiencing the pandemic. For example, the early COVID-19 crowdsourcing programme produced information on behaviours and experiences across age and gender groups, illustrating the potential value of rapid voluntary data collection when conventional survey operations were constrained.

This example also illustrates an important limitation. Voluntary crowdsourcing does not automatically produce representative estimates of the Canadian population. Participation is self-selected, and the resulting sample may differ systematically from the population.

Its value therefore lies partly in what it can contribute **alongside** official probability-based statistics and administrative data. During a crisis, it can provide rapid signals, identify questions requiring further investigation and reveal dimensions of lived experience that may subsequently inform more rigorous statistical collection.

The pandemic consequently provides a useful model for complementarity: alternative data sources can increase responsiveness without being mistaken for a replacement for representative official statistics.

9. Housing and homelessness: community data and lived experience

Housing and homelessness provide perhaps an even clearer illustration of the potential contribution of community-generated data.

Traditional statistical measures provide essential information about housing affordability, core housing need, shelter use and homelessness. Yet homelessness is a particularly difficult phenomenon to measure because some people are not visible in administrative systems or conventional surveys. Unsheltered homelessness, hidden homelessness and people moving between temporary

arrangements can be difficult to capture through periodic statistical snapshots (HICC, 2024).

Community-led counts can address part of this problem. Point-in-Time counts, for example, combine coordinated enumeration with information gathered locally about people experiencing homelessness. The nationally coordinated approach has involved 74 communities in 2024 (HICC, 2024).

An even more dynamic example is the **By-Name List**, or Unique Identifier List, used within Canada's Coordinated Access systems. A By-Name List is designed to provide a real-time picture of people experiencing homelessness who have interacted with the service system and consented to having their information collected and shared for coordinated access. The data can include demographic information, housing history, service needs and other relevant characteristics.

This produces a fundamentally different kind of statistical information from a periodic national survey. Rather than providing a static estimate at a particular point in time, community systems can monitor inflows, outflows, service needs and changes in individual circumstances.

The Canadian federal Reaching Home program has explicitly supported the development of data-driven Coordinated Access systems and By-Name Lists. Government evaluations have recognized their potential to help communities establish baselines, monitor progress and identify trends and challenges.

The complementarity is therefore particularly strong. Official statistics can provide comparability and a national framework; community data can add local context, operational relevance and a more continuous view of people's experiences.

Importantly, these data should not simply be absorbed into the official statistical system without regard to their purpose and governance. They arise from relationships between people and service providers and often depend on consent. Their value is inseparable from the trust on which the data collection relationship is based.

10. Labour markets: from newspaper advertisements to digital signals

The evolution of labour-market information provides an instructive historical example of how official statistics have long incorporated non-traditional data.

From 1962 until 2003, Statistics Canada produced a Help-Wanted Index based on job advertisements in selected newspapers. It served as an indicator of labour demand and a proxy for unmet demand. The index was eventually discontinued as employers increasingly moved job advertising to the Internet, making the newspaper-based methodology less relevant (Statistics Canada, 2003).

The historical example is revealing because the underlying statistical principle has not changed. Statistical agencies have always sought indicators that can provide timely information about economic conditions. What changes is the data environment.

Today, online job postings can provide near-real-time signals about labour demand, occupations, skills and geographic patterns. Digital platforms such as Indeed and LinkedIn generate information at a scale and frequency that traditional surveys may not reproduce.

Such information can therefore complement conventional labour-force surveys by providing more immediate signals of changing demand. At the same time, job postings are not equivalent to job vacancies, and digital-platform coverage may vary across sectors, occupations and employers. They require methodological evaluation and careful interpretation.

The lesson from the Help-Wanted Index is therefore not that official statistics should abandon conventional methods in favour of digital data. It is that statistical systems must continually adapt to changes in how economic and social activity is conducted and recorded.

11. Quality of Life: measuring conditions and understanding experience

The Quality of Life Framework offers another opportunity for complementarity because well-being has both objective and subjective dimensions.

Official statistics can measure conditions such as income, employment, housing, health, food insecurity, environmental conditions and access to services. The framework also explicitly recognizes subjective well-being through measures such as life satisfaction, meaning and purpose, and future outlook.

Citizen and community data can potentially extend this perspective by capturing perceptions, priorities and lived experiences at more local levels.

Several Canadian initiatives illustrate this broader ecosystem. Community-based initiatives such as **Vital Signs**, produced through community foundations, have used local information and community engagement to examine themes including housing, health, belonging and inclusion. Other organizations have developed surveys and engagement activities around social connection, loneliness and community well-being. Community service networks such as 211 also generate information from people seeking assistance with housing, food, mental-health and other services.

The analytical value of these sources is different from that of national indicators. They can provide early signals of emerging needs, reveal local priorities and add qualitative meaning to quantitative measures.

For example, an official indicator may establish that housing affordability is deteriorating. Community-level data may reveal which groups are seeking assistance, what barriers they face and how the problem is perceived locally. These sources should not necessarily be merged into one indicator. Instead, they can be brought into dialogue to produce a more complete understanding of the phenomenon.

This distinction is central to responsible data integration: **complementarity does not always mean physical integration of datasets. It can also mean analytical integration of different forms of evidence.**

12. Indigenous data contexts and data sovereignty

Indigenous data represents a particularly important dimension of Canada's evolving data ecosystem because questions of data quality cannot be separated from questions of jurisdiction, self-determination and governance.

The First Nations principles of OCAP®—Ownership, Control, Access and Possession—establish principles for how First Nations data and information should be collected, protected, used and shared. OCAP® is a tool supporting First Nations data sovereignty and recognizes First Nations' authority over data relating to their communities (First Nations Information Governance Centre, 2026).

This has major implications for the concept of complementarity. A community-generated dataset cannot simply be regarded as an additional resource available to the national statistical system. The community's rights and governance

arrangements are fundamental to determining whether, how and for what purposes data can be accessed and used.

This principle also reflects the limitations of treating communities merely as respondents or sources of data. Indigenous-led data initiatives can establish different priorities, concepts and definitions of well-being and can support information systems that respond directly to community needs.

Statistics Canada has increasingly recognized the importance of Indigenous statistical capacity and partnerships. Its Indigenous Statistical Capacity Development Initiative seeks to work with Indigenous organizations and communities to understand their data needs and build capacity.

The relevance to disaggregated data is twofold. First, Indigenous-led data can help address gaps in conventional statistical systems. Second, it can challenge the assumption that the objective of complementarity is necessarily to incorporate all data into a single centralized system.

A more appropriate approach may sometimes be **interoperability without centralization**: data systems can be capable of working together while communities retain meaningful control over their own data assets.

This is particularly important in the context of Truth and Reconciliation and the recognition of Indigenous rights. Data governance must support—not undermine—self-determination.

13. Data integration and record linkage

The technical foundation for many forms of complementarity is data integration. In the Canadian public sector, however, the capacity to integrate data is also an institutional and governance issue, particularly where information is distributed across federal, provincial and territorial jurisdictions (Dudoit and Labilloy, 2025; Dudoit, Labilloy and Oliveira, 2026).

Record linkage can combine information from different datasets to create richer analytical resources. Statistics Canada has extensive experience in this area, including the development of processes for assessing linkage quality, managing confidentiality and evaluating linked datasets (Statistics Canada, 2017; Statistics Canada, 2020a).

The DDAP itself uses data linkage as one mechanism for producing more detailed information. Statistics Canada can securely combine census and survey data with administrative information collected by federal, provincial and territorial organizations. These linkages can increase analytical capacity while potentially reducing the need to ask Canadians for information repeatedly.

The benefits can be substantial. Linking information can increase sample sizes, improve coverage, support longitudinal analysis and enable intersectional disaggregation.

But linkage is not simply a technical exercise. Every additional dataset introduces questions about purpose, consent, legal authority, data quality, compatibility of concepts and potential risks of re-identification.

Linkage errors can also produce statistical bias. Variables that appear to identify the same individual may differ across systems because of missing information, inconsistent formats, name changes or other discrepancies. Statistics Canada's methodological work explicitly recognizes linkage error as an important quality consideration (Statistics Canada, 2020a).

The principle should therefore be **link where there is a clear analytical and public-interest purpose, and where the benefits justify the risks and costs.**

14. Interoperability: beyond technical compatibility

Data integration also depends on interoperability. Interoperability is sometimes reduced to the ability of computer systems to exchange information. In a statistical ecosystem, however, it has a much broader meaning (Dudoit and Labilloy, 2025; Dudoit, Labilloy and Oliveira, 2026).

Datasets must be interoperable conceptually, semantically, technically and institutionally. The same term must have sufficiently compatible meanings; classifications must be aligned; metadata must be available; geographic identifiers must be compatible; legal and policy frameworks must permit appropriate exchange; and organizations must be willing and able to collaborate.

This is why statistical standards matter so much. The DDAP explicitly identifies approved statistical standards as one of its four guiding principles. The development and adoption of consistent standards make it possible to compare information across programmes and institutions.

Interoperability also extends to legislation and regulation. Different jurisdictions may have different privacy requirements, data-access rules and institutional mandates. A technically feasible linkage may nevertheless be legally or ethically inappropriate.

Finally, interoperability requires organizational culture. Data producers must be willing to collaborate, share metadata and understand each other's needs. It requires moving from a culture in which datasets are treated as institutional assets to one in which data are understood as part of a broader public information ecosystem (Dudoit and Labillois, 2025; Dudoit, Labillois and Oliveira, 2026).

In this sense, interoperability is not merely an information-technology project. It is a **collective behaviour and governance challenge**.

15. Artificial Intelligence and the Emerging Data Ecosystem

The growing use of artificial intelligence introduces another dimension to the relationship between official statistics and citizen-related data (Dudoit, Labillois and Oliveira, 2026). AI systems can process large volumes of structured and unstructured information, identify patterns across heterogeneous datasets, support classification and coding, and generate more timely analytical signals. In principle, these capabilities could strengthen the ability of statistical systems to identify emerging social and economic trends and to explore dimensions of population diversity that may be difficult to capture through conventional statistical production processes.

For disaggregated data, AI may also create opportunities to analyse information across multiple dimensions and to identify patterns that would otherwise remain difficult to detect. Natural-language processing, for example, can potentially extract information from qualitative or community-generated sources, while machine-learning techniques can support the classification and linkage of records from different sources. Such applications could complement more traditional statistical methods and help statistical organizations respond more rapidly to emerging information needs.

However, the use of AI does not eliminate the fundamental problems associated with data quality, representativeness, privacy or governance. Indeed, it can amplify them. An AI system trained on incomplete or systematically biased data may reproduce or magnify existing inequalities. Models may also generate apparently precise results without providing sufficient transparency about the underlying data,

assumptions or decision processes. This is particularly important when AI is used to produce information about populations that have historically been underrepresented or misrepresented in official statistics.

The distinction between **data quality and computational capability** is therefore critical. More powerful analytical tools do not necessarily produce more accurate or more equitable statistics. AI can process more data, more quickly, but it cannot by itself determine whether those data provide an appropriate representation of the population or whether the concepts being measured are meaningful. Human expertise, statistical standards, methodological safeguards and engagement with affected communities remain essential.

AI also reinforces the importance of interoperability (Dudoit and Labilloy, 2025; Dudoit, Labilloy and Oliveira, 2026). The value of AI applications depends in part on the ability to access, connect and interpret data across institutional and technological boundaries. Without common standards, metadata, classifications and governance arrangements, increasingly sophisticated analytical tools may simply process fragmented datasets at greater speed. Interoperability is therefore not only a technical prerequisite for data integration; it is also an important condition for trustworthy and responsible AI-enabled statistical innovation.

For official statistical systems, this suggests that AI should be understood neither as a replacement for statistical methodology nor as a substitute for trusted data institutions. Rather, it can be viewed as an additional tool within a broader ecosystem in which official statistics, administrative data, citizen-generated information and other sources interact. The central challenge is to ensure that technological innovation strengthens, rather than weakens, the principles of quality, transparency, accountability, privacy, equity and public trust that underpin official statistics.

16. Evolving statistical standards

Disaggregated data place pressure on statistical standards because concepts evolve as society changes.

The addition of sexual orientation to the 2026 Census provides a useful example. Following consultation and extensive testing, Statistics Canada introduced a question on sexual orientation for people aged 15 and older in the 2026 Census. The new information is intended to address important data gaps and support analysis of inequalities in areas including employment, housing, income and education.

The example demonstrates why standards are not merely technical classifications. Definitions determine who becomes visible in statistical systems.

Statistics Canada developed a new standard on sexual orientation in 2023 following consultation and testing. The process illustrates the importance of conceptual clarity, respondent understanding, comparability and engagement with communities when introducing new statistical concepts.

This also illustrates a key tension. Greater detail can improve visibility, but it can create privacy risks or statistical instability when populations are small. The goal is therefore not maximal disaggregation under all circumstances. It is **meaningful and responsible disaggregation**.

Statistical standards must consequently evolve alongside social realities while preserving comparability and quality.

17. Data quality and representativeness

The principal danger in expanding the data ecosystem is the assumption that more data necessarily means better data.

Citizen-generated and alternative data sources often have distinctive quality characteristics. Voluntary surveys may suffer from self-selection. Administrative data may reflect the operational requirements of a programme rather than the conceptual requirements of a statistical measure. Digital-platform data may exclude people who do not use the platform. Community data may be highly informative locally while lacking national representativeness.

The appropriate response is not to reject such sources but to characterize them accurately.

The emerging UN guidance on citizen data recognizes that there cannot be a single quality model applicable to all forms of citizen data. Instead, quality needs to be assessed in relation to the purpose, production process, coverage, engagement and intended use of the data (United Nations, 2025).

This is consistent with the broader quality framework used by statistical organizations. Relevance, accuracy, timeliness, accessibility, coherence and interpretability may involve trade-offs. A rapidly available community signal may have lower population representativeness but greater timeliness and local relevance.

The appropriate analytical question is therefore not simply:

Is this data source representative?

It is:

Representative of what, for whom, for what purpose, and under what conditions?

This more nuanced approach allows complementary sources to contribute without overstating what they can support.

18. Privacy, consent and trust

The expansion of data sources also increases the importance of privacy and trust.

Canada has strong legislative protections governing personal information, including the *Statistics Act*, the *Privacy Act* and the *Access to Information Act*. Statistics Canada emphasizes that information obtained through surveys, censuses and administrative data is protected by law.

However, legal compliance is only one component of trust (Labillois, 2026).

People need to understand why data are collected, how they will be used, who will have access and what safeguards exist. Community organizations and service providers may face an even more direct relationship with individuals whose data they collect. In homelessness services, for example, consent is integral to the use of person-specific information within coordinated-access systems.

Trust also depends on transparency. If people contribute data but never see how those data improve services or policies, the legitimacy of future collection may be weakened.

Communication therefore becomes part of data governance. Trust must operate in multiple directions: institutions must listen to communities; communities must understand institutional constraints; and both must communicate clearly with the public (Labillois, 2026).

This is particularly important for sensitive characteristics used in disaggregation. The collection of information about disability, Indigenous identity, racialized identity, sexual orientation or other characteristics can create both benefits and risks. The statistical value of such information must be balanced against the possibility of stigma, discrimination, misuse or unintended disclosure.

19. Governance and stewardship

The complexity of the data ecosystem requires governance that goes beyond technical data management (Labillois, 2026; Dudoit, Labillois and Oliveira, 2026).

Governance must establish who has authority to collect data, who can access them, for what purposes, under what conditions, and who is accountable for decisions about their use.

This is especially important when several organizations contribute to a single analytical product. A national statistical office may provide methodological leadership, while another government department owns administrative records, a community organization generates complementary information, and a private organization operates the technological infrastructure.

Clear roles and responsibilities are therefore essential.

Stewardship also implies responsibility for the full life cycle of data: collection, storage, linkage, analysis, dissemination, retention and destruction. It includes metadata, documentation, quality assessment, privacy protection and communication.

For Indigenous data, stewardship must also incorporate sovereignty and community governance. OCAP® demonstrates that governance can include collective rights and jurisdiction that differ fundamentally from conventional institutional ownership models.

A mature Canadian data ecosystem will therefore need governance arrangements capable of accommodating different forms of authority rather than imposing one model on all data.

20. Human rights, equity and inclusion

The ultimate rationale for disaggregated and complementary data is not methodological sophistication for its own sake. It is the possibility of improving the visibility of people and communities whose experiences may otherwise remain hidden.

A human-rights-based approach to data places particular emphasis on participation, self-identification, privacy, non-discrimination, equality and

accountability. OHCHR guidance explicitly encourages collaboration between national statistical offices and human-rights institutions and calls for the participation of rights-holders in data definition, collection and analysis (OHCHR, 2018).

Citizen and community data can make an important contribution because they may capture experiences that are difficult to identify through conventional statistical instruments. The United Nations has highlighted the potential of citizen-generated and community-based data to make data collection more reflective of people's lived experiences, including in the context of disability (OHCHR, 2022).

But inclusion also requires caution. Collecting information about a marginalized population does not automatically benefit that population. Data can be used to identify inequalities, but they can also reinforce stereotypes or enable harmful profiling.

The relevant test is therefore whether data collection and use contribute to **meaningful participation, equality, accountability and improved outcomes**, while minimizing potential harms.

In this sense, disaggregated data can be understood not merely as a statistical consideration, but as an infrastructure for social accountability.

21. Toward a complementary Canadian data ecosystem

The examples examined in this paper suggest a model of complementarity with several dimensions.

First, **official statistics provide the backbone**. They establish concepts, classifications, representative measures, longitudinal series, national comparability and public accountability.

Second, **administrative data provide depth and continuity**. They can offer large-scale and longitudinal information generated through programme and service delivery.

Third, **citizen and community data provide proximity**. They can identify lived experiences, local priorities and emerging needs.

Fourth, **digital and commercial data can provide timeliness**. They can generate high-frequency signals about rapidly changing economic and social conditions.

Fifth, **Indigenous and community-governed data can provide distinct perspectives and forms of authority**, challenging assumptions about who defines the questions and who controls the resulting information.

These components do not need to be merged into a single database. Indeed, in some circumstances they should not be (Di Gennaro Splendore and al., 2026). Complementarity can take the form of linkage, harmonization, parallel indicators, triangulation, methodological validation, contextual interpretation or joint analysis.

The conceptual model is therefore simple:

Official statistics + complementary data sources → richer and more granular evidence → better-informed decisions.

But the plus sign should not be interpreted as automatic addition. It represents a process of **evaluation, governance, interoperability and interpretation**.

22. Future directions

Several developments could strengthen this ecosystem.

The first is greater interoperability. Common standards, metadata, classifications and geographic identifiers can make it easier to combine or compare information.

The second is stronger capacity for intersectional analysis. Producing detailed data is only the first step. Analysts and policymakers need to acquire, maintain and apply the skills to interpret multidimensional patterns responsibly.

The third is sustained community engagement. Communities should not be approached only when a statistical organization needs respondents. They should be partners in identifying data gaps, defining concepts, assessing methods and interpreting results.

The fourth is improved data literacy. The growing abundance of data increases rather than reduces the need for methodological awareness. Decision-makers must understand not only the differences between representative estimates, administrative records, voluntary responses, community-generated evidence and digital signals, but also the limits, biases and appropriate uses of each. Improved data literacy means being able to assess whether a source is fit for the decision at hand, to recognize when uncertainty or selection effects make an inference fragile, to ask for the right complementary evidence, and to avoid treating all data as interchangeable simply because they are abundant.

The fifth is continued evolution of statistical standards. As concepts and social realities change, standards must adapt while preserving comparability.

Finally, collaboration itself needs to be treated as an institutional capability (Labillois, 2026; Dudoit and Labillois, 2025). The future statistical system will require sustained partnerships among statistical offices, governments, Indigenous organizations, municipalities, community groups, researchers and private-sector data producers.

The emerging international work on citizen data points in this direction. The Copenhagen Framework seeks to establish an enabling environment in which citizens, communities and civil society organizations can contribute to national data ecosystems while maintaining appropriate principles and safeguards (United Nations, 2025).

Canada is therefore participating in a broader international transition from a model centred principally on institutional data production toward a more pluralistic data ecosystem.

23. Conclusion

Disaggregated data are essential if statistical systems are to reflect the diversity of contemporary societies. Aggregate statistics remain indispensable, but averages alone cannot reveal all forms of inequality, exclusion or difference.

Canada's Disaggregated Data Action Plan represents an important institutional response. Its emphasis on disaggregation, intersectionality, statistical standards and geography provides a framework for moving toward more detailed and inclusive official statistics. The Quality of Life Framework similarly demonstrates the value of broadening the statistical portrait beyond conventional economic measures and incorporating both objective and subjective dimensions of well-being.

At the same time, official statistics cannot—and should not be expected to—capture every dimension of a rapidly changing society through conventional collection mechanisms alone. Administrative records, digital interactions, community-generated information, voluntary surveys and other citizen-related data can provide complementary perspectives.

The examples considered here illustrate different forms of value. COVID-19 crowdsourcing demonstrated the value of rapid information under extraordinary circumstances. Homelessness By-Name Lists demonstrate the potential of

continuous community-level information. Digital labour-market data illustrate how new economic behaviours generate new statistical signals. Quality of Life initiatives show how official indicators can be complemented by information about perceptions and lived experience. Indigenous data governance demonstrates that complementarity must sometimes operate within principles of sovereignty and community control.

The central challenge is therefore not whether Canada should choose official statistics or citizen data. **It is how the two can coexist within a trusted, interoperable and rights-respecting data ecosystem.**

That requires careful attention to quality, representativeness, privacy, consent, standards and governance. It also requires a cultural shift: from viewing data primarily as institutional assets toward recognizing their role as part of a broader social and economic infrastructure (Dudoit, Labillois and Oliveira, 2026; Labillois, 2026).

Ultimately, disaggregated data are not simply about producing more numbers on smaller groups or more detailed tables. They are about making people and their experiences visible. Citizen and community data can help strengthen that visibility—but only when their limitations, rights implications and governance requirements are taken as seriously as their potential.

The objective should therefore not be to replace official statistics, but to make the statistical system **more inclusive, more responsive and more capable of describing the complexity of the society it serves.**

References

Crenshaw, K. (1989). Demarginalizing the intersection of race and sex: A Black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics. *University of Chicago Legal Forum*, 1989(1), 139–167.

Di Gennaro Splendore, L. (2024). Is there a quantitative relationship between democracy and official statistics? *Statistical Journal of the IAOS*, 40, 123–141.

Di Gennaro Splendore, L., Labillois, T., Camilleri, C., Macfeely, S. and Vassallo, M.T. (2026). Federalism and the quality of official statistics: Comparative lessons from Argentina, Canada, Germany, India, Nigeria, Switzerland and the United States. *Statistical Journal of the IAOS*. OnlineFirst, 17 June 2026. <https://doi.org/10.1177/18747655261455431>.

Dudoit, A. and Labillois, T. (2025). *Digital Sovereignty and federalism: Data Interoperability and AI governance*. CIRANO, 2025PR-12.
<https://doi.org/10.54932/DNAN8237>.

Dudoit, A., Labillois, T. and Oliveira, C. (2026). *Public Sector Data Interoperability and Responsible Adoption of AI in Canada: Strategic Overview and Call to Action*. CIRANO, 2026PR-02. <https://doi.org/10.54932/PUUN9151>.

First Nations Information Governance Centre (2019). *First Nations data sovereignty in Canada*. Statistical Journal of the IAOS, 35(1), 1–13.

First Nations Information Governance Centre (2026). *The First Nations Principles of OCAP®*. Ottawa: First Nations Information Governance Centre.

Government of Canada (2021). *Budget 2021: A Recovery Plan for Jobs, Growth and Resilience*. Ottawa: Government of Canada.

Government of Canada (2023). *Statement and Impacts Report on Gender, Diversity, and Quality of Life*. Ottawa: Government of Canada.

Government of Canada (2024). *Budget 2024: Safer, Healthier Communities*. Ottawa: Government of Canada.

Government of Canada (2025). *2025–26 Departmental Plan*. Ottawa: Government of Canada.

Hicks, P. (2022). The System of National Quality-of-Life Statistics: Conceptual Framework. *Analytical Studies: Methods and References*, Statistics Canada, Catalogue no. 11-633-X.

Housing, Infrastructure and Communities Canada (2024). *Everyone Counts 2024 – Highlights Report Part 1: Enumeration of Homelessness*. Government of Canada source

Housing, Infrastructure and Communities Canada. *Reaching Home Directives*. Government of Canada source

Labillois, T. (2026). *Leadership as Stewardship: Cultivating Ethical Culture and Communicating Values in National Statistical Offices*. Expert Meeting on Ethics in Official Statistics, United Nations Economic Commission for Europe (UNECE), Geneva, 12–13 March 2026.

Office of the United Nations High Commissioner for Human Rights (OHCHR) (2018). *A Human Rights-Based Approach to Data: Leaving No One Behind in the 2030 Agenda for Sustainable Development*. Geneva: United Nations.

Office of the United Nations High Commissioner for Human Rights (OHCHR) (2022). *Guidelines on Deinstitutionalization, Including in Emergencies*. Geneva: United Nations.

Statistics Canada (2003). *Help-wanted Index, April 2003*. The Daily, 6 May 2003. Ottawa: Statistics Canada.

Statistics Canada (2017). *Record Linkage Project Process Model*. Ottawa: Statistics Canada, Catalogue no. 12-605-X.

Statistics Canada (2020a). *Statistical Methodology Research and Development Program Achievements, 2019/2020*. Ottawa: Statistics Canada, Catalogue no. 12-206-X.

Statistics Canada (2020b). *2020–21 Departmental Results Report*. Ottawa: Statistics Canada.

Statistics Canada (2021). *Moving Forward on Well-being (Quality of Life) Measures in Canada*. Analytical Studies Branch Research Paper Series. Ottawa: Statistics Canada.

Statistics Canada (2023a). *Statistics Canada's Disaggregated Data Action Plan Achievements Report 2022–23: Building on a Solid Foundation*. Ottawa: Statistics Canada.

Statistics Canada (2023b). *Statistical Methodology Research and Development Program Achievements, 2022/2023*. Ottawa: Statistics Canada, Catalogue no. 12-206-X.

Statistics Canada (2024a). *2023–24 Departmental Results Report*. Ottawa: Statistics Canada.

Statistics Canada (2024b). *Quality Assurance Framework*. Ottawa: Statistics Canada.

Statistics Canada (2025a). *Evolving our approach to measuring quality of life in Canada*. Ottawa: Statistics Canada.

Statistics Canada (2025b). *Statistics Canada's 2024–25 Departmental Results Report*. Ottawa: Statistics Canada.

Statistics Canada (2025c). *Disaggregated Data Action Plan Accomplishments Report 2024–2025: From Commitment to Continuity*. Ottawa: Statistics Canada.

Statistics Canada (2025d). *Report on Content Changes for the 2026 Census of Population: Sexual Orientation*. Ottawa: Statistics Canada, Catalogue no. 98-20-0005.

Statistics Canada (2026a). *Disaggregated Data Action Plan*. Ottawa: Statistics Canada.

Statistics Canada (2026b). *Quality of Life Framework for Canada*. Ottawa: Statistics Canada.

Statistics Canada (2026c). *Framework for the Use, Analysis, Interpretation and Dissemination of Police-reported Indigenous and Racialized Identity Data Through the Uniform Crime Reporting Survey*. Ottawa: Statistics Canada, Catalogue no. 11-633-X.

Statistics Canada (2026d). *Help-wanted Index, monthly*. Table 14-10-0280-01. Ottawa: Statistics Canada.

United Nations (2014). *Fundamental Principles of Official Statistics*. United Nations General Assembly Resolution A/RES/68/261. New York: United Nations.

United Nations (2025). *Copenhagen Framework on Citizen Data*. United Nations Statistics Division, Collaborative on Citizen Data. New York: United Nations.

United Nations Statistics Division (2025). *Guidance Reports to Support the Use and Improvement of Citizen Data*. New York: United Nations.