Pathways Towards Data Interoperability for the Sustainable Development Goals in Canada

April 2019

MaRS Possibilian Carleton University IISD La fondation McConnell Foundation

Executive Summary

In order for Canada to play its role in achieving the United Nations Sustainable Development Goals (SDGs), organizations across all sectors need access to more and higher quality data to inform decision-making and evaluate progress towards the SDGs. This requires investing not just in collecting data, but also in joining up existing and emerging data sets. When data sets from different sources can be accessed, processed, and integrated without losing meaning, data becomes interoperable, which in turn unlocks massive network effects. Interoperability could increase data sharing, data quality, and automated data processing, while reducing fragmentation of data and resource intensity of data collection, creating significant new value for Canadian organizations and society.

This report maps the current landscape of SDG data interoperability in Canada, synthesizing participatory workshops and interviews with a survey of the literature and online sources. We articulate a desirable future state - the Internet of Impact - where more interoperable data accelerates progress towards the SDGs. We then chart plausible pathways towards the desired future, considering syntactic and semantic interoperability; human and machine data integration; and centralized and distributed approaches to interoperability. Interoperability is not a single problem with a single solution; but it is a necessary investment to unlock the power of data for good. By leading the development of SDG data interoperability, Canada can accelerate its own contributions towards the SDGs, while also making a valuable contribution to a difficult problem that for the first time is within reach of being solved.

Table of Contents

INTRODUCTION 4
THE CASE FOR SDG DATA INTEROPERABILITY 5
CURRENT STATE: WHAT WE KNOW 8
FUTURE STATE: WHAT WE WANT 15
PATHWAYS TOWARDS INTEROPERABILITY 16
SYNTACTIC VS SEMANTIC INTEROPERABILITY 17
HUMANS VS MACHINES 20
CENTRALIZED VS DISTRIBUTED DATA 21
THE CANADIAN SDG DATA ECOSYSTEM 22
PROBLEMS THAT DATA INTEROPERABILITY DOESN'T SOLVE 24
CONCLUSION: A LEADERSHIP OPPORTUNITY FOR CANADA 25
ACKNOWLEDGEMENTS 26
REFERENCES 27
ANNEXES 28
ONTOLOGIES FOR THE REPRESENTATION OF UN SDG INDICATORS 29
SDG INTEROPERABILITY INTERVIEW GUIDE 35
SDG INTEROPERABILITY PARTICIPATORY WORKSHOP METHODOLOGY 36
SDG INTEROPERABILITY LITERATURE SUMMARY 36

Introduction

There is a growing recognition that sustainability is at the heart of everything we do. It is now at the heart of business operations, product design, policy creation, investment risk, and is driving innovation. The 2030 Agenda for Sustainable Development, adopted by United Nations Member States in 2015, provides for the first time, a global, shared blueprint for peace and prosperity, for people and the planet, now and into the future. The 2030 Agenda provides a comprehensive set of 17 Sustainable Development Goals (SDGs), supported by 169 targets and 230 indicators to measure progress towards achieving the goals.

Increased focus on sustainable development has brought increased focused on sustainable data: What new data should be gathered? What existing data can be mobilized to tell a sustainability story? How can data be shared, aggregated and combined in novel ways to improve sustainable development?

Data is increasingly abundant. It's created and captured in nearly every part of society and daily life. Whether it is from our mobile devices, internet usage, sensors and satellites, or surveys and process information, there is data everywhere and it is growing at an accelerating rate. A key challenge in sustainable development is not only to identify and fill data gaps, but to mobilize existing data to inform decision making and assess impact.

While data is everywhere, accessing that data is difficult. Even when data is useful to sustainable development, the data was likely collected for other reasons, and usually as a byproduct of other activities. Data owners are often not sufficiently motivated or resourced to be able to facilitate open access to the data. Accessing that data then requires permission, the ability to access and receive the data, and finally, the ability to use that data to produce information that is useful for sustainability.

Ocean Protocol (https://oceanprotocol.com/) is an example of a new type of data ecosystem allowing data providers and users to share, use, and extract value from data in new ways. These new technologies and the communities forming around them point to a future where data ownership, access, and reuse will be more common and dynamic. In the interim, however, data assets are largely stranded, undervalued, and when available, difficult to combine with other data to create information that is needed to understand, evaluate and enhance impact. This is proving to be a particular challenge to the SDGs, which are often interconnected and meant to be measured on systemic and national scales.

When government and public agencies undervalue and under-invest in data, the public interest is put at a disadvantage. Private sector organizations on the other hand value their data as a core asset and are increasingly turning to social issues where they identify market opportunities. Apple, Amazon, Facebook, and Google, for example, are leveraging their data to make commercial entries into healthcare, smart cities, and other domains that have traditionally been considered the realm of government. Because the private sector investment in data and analytics dwarfs public sector investment, the value derived from analytic insights is also being captured privately. In many cases, ownership of data accrued through government services already resides with third parties who provide technical or service delivery. Many civil society organizations have opportunities to collect useful data, but lack the resources to capitalize on these opportunities. Until data is valued and resourced as an asset in the public interest, a government or public interest institution's ability to use data to advance impact will be increasingly constrained.

The ability to access and make use of existing and emerging data is served by two parallel paths: first, supporting the development of data ecosystems around impact ecosystems; and second, supporting the advancement and adoption of standards and interoperability of sustainable development data. In the context of the current undervaluing and under-resourcing of data in the public interest, there is an urgent and catalytic opportunity to provide the capacity to pool, grow, and coordinate resources towards these aims on a national and global scale.

The purpose of this report is to identify a path forward for Canada towards interoperability of sustainable development data. We begin by articulating the case for investing in improved interoperability. Adopting the SDGs as a global framework for measuring sustainable development, we map the current landscape of SDG impact data interoperability in Canada, developing eight use cases and presenting the results of a global literature review, national and international subject matter expert interviews, and three virtual participatory community workshops. Synthesizing the needs and interests of workshop and interview participants, we describe a desirable future state where more interoperable data accelerates progress towards the SDGs. We then chart plausible pathways towards the desired future, considering syntactic and semantic interoperability; human and machine data integration; and centralized and distributed approaches to interoperability. Finally, we recognize the limitations of improved data interoperability, documenting SDG data challenges that will require advances in other areas.

The Case for SDG Data Interoperability

What is Data Interoperability?

The Global Partnership for Sustainable Development Data provides a useful definition of interoperability:

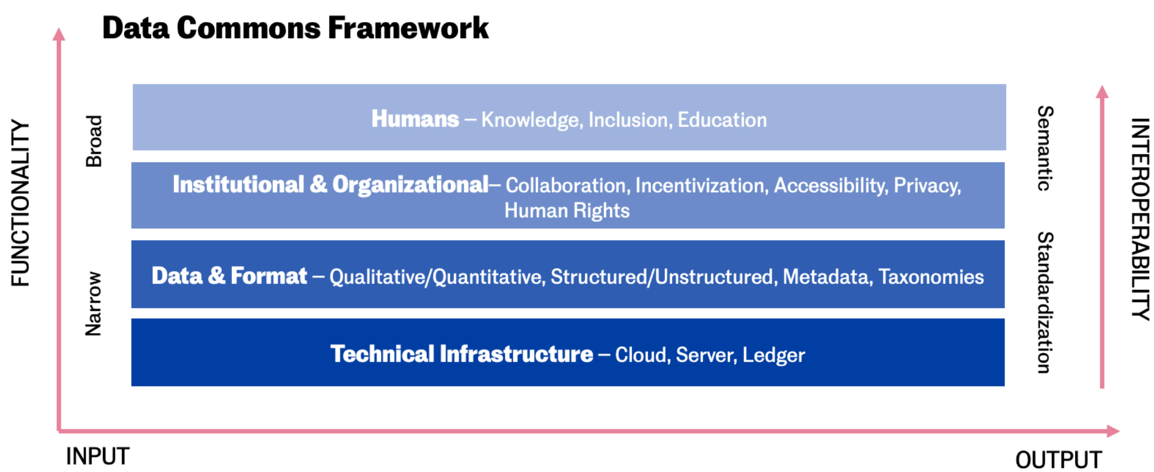
> “Interoperability is the ability to access and process data from multiple sources without losing meaning and then integrate that data for mapping, visualization, and other forms of representation and analysis. Interoperability enables people to find, explore, and understand the structure and content of data sets. In essence, it is the ability to ‘join-up’ data from different sources to help create more holistic and contextual information for simpler, and sometimes automated analysis, better decision-making, and accountability purposes.”[1]

Joined-up data requires many layers of interoperability. It is more than just compatible formats for data sets. There also must be consistent and agreed meanings of words inside the datasets. Without agreement on the meaning - or semantics - of data, the conclusions drawn from aggregated data may be suspect at best and incorrect at worst. Thus, while it may be tempting to think of interoperability as a technical problem to be solved by computer scientists, in fact it is a multi-dimensional problem that requires a range of expertise.

Goldstein’s Data Commons Framework[2] has been summarized into four layers – technology, data and format, institutional and organizational, and human – by the The Global Partnership for Sustainable Development Data:

- The technology layer addresses the standards needed to make data accessible on the Internet;
- The data and format layer focuses on data structures, metadata standards, and vocabularies;
- The institutional and organizational layer covers process standards needed to keep data accurate and consistent, as well as high-level policies such as data sharing agreements ; and
- The human layer emphasizes the need for common understandings among those who produce and use the data.

Together, these layers show the elements that are needed for successful data interoperability.



Why should Canada want better Data Interoperability?

With better interoperability of impact data, Canada can expect the following benefits:

- Greater sharing of data across organizations and sectors;
- Greater consistency, reliability, intelligibility, and useability of data;
- Greater automation of data processing;
- Reduction in fragmentation and duplication of data;
- Reduction in time and resources allocated to collecting data; and
- Greater value from data-driven insights for decision makers at all levels.

When data interoperability is translated into practice:

- Relationships between entities producing, controlling and using data are improved;
- Public, private, and civil society organizations that produce data are more open; and
- Existing data inventories are modernized and are made more accessible and useable.

The objective is not to maximize interoperability, but to strike a balance between the benefits of interoperability and legitimate concerns around privacy, security, and potential misuses of data. At present, Canada has much to gain from greater interoperability.

If data were more interoperable...

"...I could make more compelling conclusions, highlight and underline successes and challenges more clearly, and get our data onto our site and into the public's hands significantly more quickly."

– **Heather Block,** *United Way Winnipeg*

Why use the SDGs to drive data interoperability?

The 17 goals, 169 targets and 230 indicators in the 2030 Agenda for Sustainable Development provide a globally recognized framework that can guide both the content and format of data. It makes sense to build upon this global consensus forged by the United Nations about the world we all want to live in. The principles of the 2030 Agenda call for multi-stakeholder partnerships for mobilizing and sharing knowledge. The principles also recognize the interdependence and indivisibility of the SDGs - success requires progress on all SDGs everywhere. Together, these two principles entail that achieving the SDGs will require a dramatic increase in the number of data interactions across entities, sectors, countries and levels of government.

The SDGs are simultaneously a driver for interoperability, and a universal framework within which interoperability can be advanced. Achieving the SDGs is increasingly a priority for organizations and communities across Canada. Data interoperability is at the heart of our ability to use data effectively and meaningfully to accelerate progress on the SDGs.

Current State: What We Know

The Beneficiaries of Interoperability

To ground our work in the specific needs of the beneficiaries of improved sustainable development data interoperability, we developed nine use cases. Each use case identifies a user segment, their roles and responsibilities, potential benefits and user-specific needs for interoperability.

The current lack of data interoperability is a multi-dimensional problem that requires a range of expertise to solve. If every data entity approaches data interoperability as a silo within special interest topics (such as increasing data collection efforts, or aligning standards within use cases), this would lead to more use-specific standards. However, this type of narrow approach risks generating insufficient value for the ecosystem to be sustained. Allowing disparate silos to continue efforts will result in increased difficulty in bringing efforts together.

To achieve interoperability for meaningful data collection and actionable insights, there must be coordinated and collective direction. We believe scalable and sustainable results will only be achieved by addressing the fundamental layers of a common approach and semantic interoperability.

1. Statistical Offices

Example: National or provincial statistical offices (e.g. Statistics Canada)

Purpose: Develop a Canadian indicator system to monitor the implementation of the Canadian SDG strategy (StatCan)

Key Challenges:

- Integrating data from sources for which StatCan did not control the methods of data collection or analysis
- Transitioning from a siloed system (internally interoperable) to an open data system (externally interoperable)

Needs from Data Interoperability:

- Access to data from other government departments and provincial sources
- Access to data from lesser quality sources that have been collected using different standards

Outcome from Data Interoperability:

- Reduce the cost and time resources required when integrating external data
- Would then generate higher value from the data after integration

2. Audit Institutions

Examples: Supreme audit organizations (e.g. Office of the Auditor General); Provincial auditors; Municipal auditors; Private sector auditors (e.g. PwC, Deloitte, KPMG)

Purpose: Assess the performance of governments and the private sector in implementing the SDGs

Key Challenges:

- Won't invest in interoperability due to narrow scope of auditing task
- Finding the right data and making it usable
- Discovering and accessing data from different sources

Needs from Data Interoperability:

- Comparing in-house data produced by clients with independent sources
- Verifying data quality from non-traditional sources
- Joining up data across levels and scales

Outcome from Data Interoperability:

- Pool resources to ease workload
- More data would be usable for auditing
- Create "audit value" from existing data
- Provide a more complete picture of client performance and establish transparency
- Coordinated approaches could increase comparability of performance

3. Governments

Examples: Government of Canada; Provincial/territorial governments; Municipal/regional governments

Purpose: The 2030 Agenda encourages governments at all levels to develop their own targets for SDG implementation

Key Challenges:

- Collecting and accessing data
- Using data collected by NSOs and other data providers for specific purposes
- Include non-traditional data sources that do not adhere to existing standards and may be inconsistent
- Bottom-up approach to SDG implementation leading to parallel reporting efforts
- Shifting from top-down data collection approach to leveraging datasets from the bottom-up

Needs from Data Interoperability:

- Every strategy laying out specific targets must be accompanied by an indicator system to monitor progress in an inclusive way
- Developing of inclusive metrics based on integrated/aggregated data
- Consistency across geographies
- Aggregating across levels of government
- Benchmark across jurisdictions

Outcome from Data Interoperability:

- Benchmark against other jurisdictions
- Support the development of more inclusive indicators that consider the needs of vulnerable groups
- Enable the joining up of data from many different sources around specific problems
- Coordinate investments in interoperability (national data strategy or framework)

4. Civil Society Organizations (CSOs)

Examples: Internationally focused organizations (e.g. Canadian Council for International Cooperation); National and provincial advocacy groups (e.g. Policy Wise, Environmental Defense Fund, Canada Without Poverty)

Purpose: Provide regular assessments on national and global issues and become stewards of comparable performance data on the SDGs

Key Challenges:

- Efforts exist in isolation from other data ecosystems
- "Incomplete" data capacity (may have data, but lack resources to use it and make it available to others)
- Influence in poor countries is limited as they depend on voluntary funding and volunteer staff

Needs from Data Interoperability:

- Understand their own organization's contribution to progress toward the goals
- Ability to aggregate local chapters to understand the impact of the whole organization
- Build a compelling evidence-based case for policy change
- Develop data for existing gaps or new problem framings (WWF's - Living planet index, Transparency international's global corruption perception index, SDSN's SDG cities index)

Outcome from Data Interoperability:

- Access and process more data sources at lower cost
- Provide more coherent and powerful messages by linking their own efforts to other data ecosystems on similar issues
- Link data from local chapters to national and international bodies
- Benchmark progress of different regions

5. Businesses

Examples: Large corporates (e.g. IKEA, Unilever, Danone, RBC); SMEs using GRI and B-Corp frameworks

Purpose: Sustainability reporting is becoming part of shareholder and financial risk assessment and is increasingly being linked to the SDG framework

Key Challenges:

- Access to data of suitable quality is not readily available to show impact
- Misuse of data by media and the public resulting in reputational damage
- Data safety and privacy to protect sensitive corporate and personal information

Needs from Data Interoperability:

- Simplified reporting to impact and responsible investors; inform eco- and socially-motivated consumers
- Compelling data for business analytics to understand business opportunities aligned with SDGs

Outcome from Data Interoperability:

- Access and process more data sources at lower cost
- Provide more coherent and powerful messages by linking their own efforts to other data ecosystems on similar issues
- Link data from local chapters to national and international bodies
- Benchmark progress of different regions

6. Cities, Municipalities, and Local Communities

Examples: Milan urban food policy pact; Global Covenant of Mayors for Climate and Energy; Youth Climate Lab

Purpose: Bottom-up principle of the 2030 Agenda empowers cities and local communities to become key actors in SDG implementation

Key Challenges:

- Lack of expertise and capacity from local communities to measure impact effectively

Needs from Data Interoperability:

- Develop locally relevant data sets
- Develop comparable data sets and indicators allowing them to benchmark performance against shared global goals and share success stories and good practices
- Smaller communities new to measurement need to access relevant data and develop available data sources

Outcome from Data Interoperability:

- Localized SDG metrics
- Allows communities to mobilize local data sources for reporting and comparison.
- Increased accessibility of data could reduce cost of data analytics
- Allows communities to understand their own contribution to national targets and shared global goals
- Enables bottom up measurement as an essential tool for bottom-up implementation
- Empowers/motivates local actors to engage through short feedback loops (citizens can see their impact)

7. Citizens

Examples: Individual actions from citizens

Purpose: The SDGs raise awareness of the impact of individual behaviour

Key Challenges:

- Unable to track individual impact and contribution to larger goals
- Lack of understanding of SDGs
- Anonymity of aggregated data sets

Needs from Data Interoperability:

- Increased transparency and accessibility of data as an education tool
- Create a sense of belonging, participation, and ability to impact

Outcome from Data Interoperability:

- More citizens are interested in understanding and tracking their personal impact on the environment and demonstrating the impact of changes in behavior

8. Research and Academia

Examples: Universities and think tanks

Purpose: Use data to develop theory and push the frontiers of knowledge

Key Challenges:

- Data collection and investments into quality stop when the funding ends
- Unrealized potential to improve data quality and useability in areas not relevant for research (unable to scale, ability to develop new uses of existing data, ability to define data needs for new problems)

Needs from Data Interoperability:

- Ability to combine and aggregate data from many datasets

Outcome from Data Interoperability:

- New insights from new combinations of data
- New ways of generating and using data
- Data innovation

What we heard in Canada

To understand the current state and needs from the sustainable development community, we engaged in multiple participatory approaches including national and international subject matter expert interviews, and three virtual community workshops involving participants from across Canada and across the use cases. A total of 27 people participated in our engagement process (the questions we asked are listed in Annex 2: SDG Interoperability Interview Guide and Annex 3: SDG Interoperability Participatory Workshops Methodology). While the representation across the sectors, organizations, and roles were diverse, similar views about data interoperability were echoed across all participants: Data users and data creators alike described common needs and common barriers to collection, access and meaningful use of SDG data.

We heard that Canadian organizations and communities desire standardized, coordinated and scalable data that can be compared across organizations, benchmarked in and across industries, and collated nationally and internationally. Although organizations and communities desire a higher level of rigour in order to ensure decision making is based on trustworthy data, they face significant capacity (both time and expertise) and resource constraints that force careful trade offs around quality and cost of data. These concerns were especially acute for small enterprises. Assessing and mitigating the privacy and security risks from collecting and sharing data adds significant costs for data creators. Given these costs and constraints, participants expressed concern about the burden of collecting and maintaining interoperable data.

> "I need data on impact that is comparable against business size/geography/industry"
>
> – **Workshop Participant,** *Impact Certification Organization*

> "Models are needed that enable small in-house teams to analyze large streams of data, ones that can also identify when reported (or missing) data is material or not."
>
> – **Kate Murray,** *Rally Assets*

In addition to the cost of sustaining interoperability, some participants questioned the relevance of SDG data. Organizations, investors, and governments invest in data to assess the success and social outcomes of their projects and investments. In cases where the organization was explicitly working toward an SDG target, they tracked SDG indicators. In all other cases, the SDGs are currently used more as a guiding framework. A common refrain, particularly from community organizations, was that the SDG indicator framework was not aligned with their organization's data needs. This points to the need for further efforts in localization of the SDGs to increase their perceived relevance to companies and communities.

Participants questioned whether there was comprehensive representation from the SDG ecosystem, and if there was motivation and willingness to participate in interoperability to generate meaningful and actionable data. They expressed a desire for continued community engagement and partnerships, combined with international alignment that would facilitate learning.

Participants expressed a lack of information that would unlock better interoperability. For example, they do not know who is doing what to make progress against which SDG goals and indicators. They do not know which funders are supporting which organizations towards what outcomes. They do not have access to future-oriented data about the plans of other organizations. They do not have access to benchmarks to compare their organization's or community's performance against their peers.

The absence of basic information results in a lack of coordination. For example, several participants observed a lack of coordination between different funders who request different data in different formats, which increases the data collection burden on community organizations without adding any real value. Participants suggested that these issues could be ameliorated, to some degree, through a coordinating body to establish consistency over time. Participants expressed optimism that improving interoperability presently could alleviate their data collection burden in the long run.

"There's a need for funders to change the way they operate and start issuing more grants for testing, monitoring, evaluation and data collection. Most funders provide financial support for programs but not the data infrastructure around them. This approach presents a barrier to change, as most organizations are limited in their capacity to build or improve data collection processes without direct funding to support those activities"

– **Gina Babinec,** *Impact Hub Ottawa*

What the Global SDG Community is saying

To complement the primary research performed through interviews and workshops, we conducted a literature review and online scan (see Annex 4: SDG Data Interoperability Literature Summary). The current global conversations about SDG Data Interoperability focus primarily on three related topics.

1. Data accessibility

There is a tremendous effort to move data onto portals and repositories to make it easy for people to find. This effort is simply seeking to surface the data that exists but is not accessible. It precedes joined-up data.

In some ways, data availability is about data interoperability. To migrate data to portals and repositories, the data must be in appropriate file formats. Some formats allow users to view the data (PDF, JPG, PNG, etc.), while others allow users to analyze and visualize the data (XSL, XLSX, CSV, TXT and JSON). The latter is referred to as machine-readable formats.

Metadata standards such as DCAT give transparency to who created the data, when it was created, where it was created and so on. SDMX is a standard that can facilitate the movement and exchange of data into common location.[3] These standards are elements of syntactic interoperability (discussed below). They are recognized as best practice of SDG data platforms.[4]

However, interoperability technologies mentioned above already exist and most are widely adopted. Interoperability itself in not the barrier. Data availability is less an interoperability problem, than it is a challenge to get useable data public.

2. Open data; private data; big data

Government open data movements are a significant piece of the SDG data availability conversation. Open data purposefully releases data in any number of formats for free and non-exclusive use by any party. Usability of data released via an open data approach is primarily limited by the skills and ability of the data user. Open data strategies render data free to be used, re-used, and redistributed. Open data initiatives, such as the Open Government Partnership, International Open Data Charter, and Open Data Institute are ensuring that the open data is kept at the forefront of national statistics.

A related effort on Data Collaboratives[5] seeks to unlock and combine data from governments and private companies to glean new insights. The data is not necessarily open to the public, but rather made available to specific users (e.g., researchers, nonprofits) in specific ways that limit deaggregation or open sharing for specific purposes, such as research.

While there are technical data interoperability issues at play, the key challenges pertain to agreements and data governance-structures that unlock the data.

3. Data governance

Data stewardship refers to the governance systems required to make private data public with limitations to access, privacy, security and control. For example, the UN Development Group has published guidelines for big data that covers topics like consent and data retention policies.[6] There are similar initiatives in Canada, such as the Tri-Statement Agency of Principles of Digital Data Management.[7] The First Nations Information Governance Centre has developed the OCAP principles for the governance of data on First Nations.[8] Because these bodies already deal with formats, standards, and contracts, they are increasingly being asked to expand their scope from data availability to enabling joined-up data.

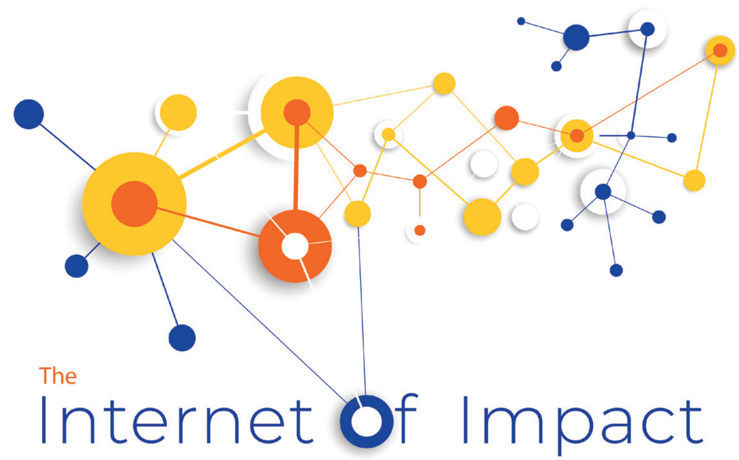
"There is an overall lack of awareness in terms of where to get the right data and lack of openness from organizations who can provide useful impact data."
– Shannon Kindornay, *Canadian Council for International Cooperation*

Future State: What We Want

The transformative possibilities of SDG data interoperability come not from simply making data available and open, but linking and joining up that data to generate game-changing insights.

The term "Internet of Impact"[9] expresses the idea of joined-up, searchable and scalable data. It imagines distributed hubs of data, that can be securely and easily shared and cited through linked files. For example, to compile data on a city, one could link to several national datasets and the data would automatically populate, and update when the original is updated. Or, to compile and analyse information on national traffic emissions, data from several cities could be seamlessly linked. Unlike existing portals that allow users to click on the links to see related datasets, the Internet of Impact would pull data from linked sites and embed it into a meaningful data table or visualization. The data could easily be accessed, aggregated and analyzed while preserving the meaning and integrity of diverse data sources.

The Internet of Impact is also an active virtual community. A coordinating body sets minimum specification standards to establish consistency and maintain quality. Everyone can see which organizations are working towards what outcomes, with visibility on both past impact and future plans. Individuals and organizations can access relevant benchmarks, connect with funders and peers, ladder up their work into higher goals, and localize the SDGs to their local community. The Internet of Impact offers multiple channels for participation and continuous engagement. It helps to align local, national, international, and global efforts towards SDG data interoperability.



Pathways Towards Interoperability

Interoperability is not solely a technical issue, but also a complex of interwoven human, organizational, legal, and social issues. Due to these interwoven layers, interoperability is inherently complex. Moreover, the SDGs themselves are indivisible and interconnected. Thus, the nature of the data around which interoperability is sought poses additional demands around breadth, depth, and interconnectedness. On the pathway towards data interoperability for SDGs, there are three major areas of consideration:

- Syntactic vs semantic interoperability
- Human-enabled vs machine-enabled interoperability
- Centralized vs distributed data architecture

Below we discuss these different options for interoperability, what they would allow Canada to accomplish with SDG data, and suggestions for how Canada can derive the most advantages from interoperability while minimizing the risks.

Syntactic vs Semantic Interoperability

Interoperability can be different things. It can simply mean that the file formats of the data can be read by different types of software and on different computers. This is called syntactic interoperability. It is easier to achieve and is sufficient for many types of information sharing. As noted above, data availability on data portals is made possible by widely available standards for syntactic interoperability.

Interoperability can also mean that any dataset can be readily combined with another dataset on the same topic. To achieve this, the words and numbers within the data set must have the same meaning. This is called semantic interoperability. Semantic interoperability deals with content, like words. A semantic interoperability problem occurs when datasets use the same words but do not refer to the same thing, and conversely, when data sets refer to the same thing with different words. Semantic interoperability has varying degrees of complexity based on what the data represents.

Complete semantic interoperability is necessary if datasets are to be combined and analyzed by machines. Weaker forms of semantic interoperability can be created to enable humans to combine data sets (which machines or humans can then analyze). The advantage of greater semantic interoperability is larger and more automated data analysis. It is much more difficult to achieve than syntactic interoperability, and perhaps not desirable to achieve completely.

Syntactic interoperability:

dataset formats and dataset labels (metadata) allow for data exchange

Semantic interoperability:

the words and numbers in the dataset have common meanings, so that a machine knows when things are the same

Syntactic Interoperability:

As noted above, syntactic interoperability is necessary to improve data availability. With good syntactic interoperability, data can be viewed, analyzed, visualized and downloaded by different computers, and it is relatively easy.

This is not to say there are not challenges. For example, one difficult and emerging area of work is version control. After data is downloaded, users must take care to incorporate all corrections, updates and improvements that are made to the original dataset, otherwise erroneous or out-of-date data circulates. With each new version, users must recommence the laborious task of cleaning and adapting the data for their purpose. This is costly, time consuming, error prone and auditing can be difficult if old versions of data sets are removed from the portal. Solving this problem requires sustaining links between the original data and the downloaded copy. There are services, such as QRI[10], working to automate repetitive tasks, ensure that the data stays in sync, and keep a history of changes so that the data can be audited even if the data owner removes older versions from websites. This is an example of ongoing innovation to improve the syntactic interoperability of available data.

However, the real power of SDG data will come not from making data available, but from joined-up or linked data. This requires additional degrees of interoperability: it moves from syntactic interoperability into semantic interoperability.

Syntactic vs Semantic Interoperability

Semantic interoperability for uniform things

The easy part of semantic interoperability is the data elements that already have unique identifying codes. For example, one can have a high degree of confidence that the Canadian Postal code “T0L 0B6” in one database refers to the same geographic catchments as “T0L 0B6” in another database. Items that already have a unique code, like postal codes, latitude and longitude, social insurance numbers, and vehicle identification numbers, can readily be linked. It is relatively easy for data creators to make their datasets more interoperable by using internationally recognized date formats, address formats and location identifiers. This semantic interoperability allows data pertaining to a particular postal code in one data set to be combined with data pertaining to a particular postal code in another data set to create a combined, richer data set. There are numerous taxonomies and vocabularies available that data creators can use to assign widely recognized codes to uniform and standardized things. Canada could choose from among these and promote one as the Canadian standard. Alternatively, Canada could simply encourage data creators to use any of the available standards and use metadata to inform others which standard was used. Ultimately, the information that is uniform in real life, is relatively straightforward to make semantically interoperabile in datasets.

Semantic interoperability for concepts with diverse and contested meanings

Semantic interoperability becomes much more fraught for conceptual ideas with diverse and often contested meanings in real life. A dataset with a column labelled “poverty” for example could refer to poverty conceptualized and measured in a number of different ways. In Canada, we might see poverty measured as a fixed income threshold, a flexible income threshold based on a market basket measure, or a proxy measure such as ‘on Ontario disability support’. These are all accepted measures of poverty, but it is unclear if an Albertan dataset using one measure can be combined with an Ontario dataset using another measure to examine correlations between nutrition and poverty. (Nutrition is another example of terms with diverse and contested meanings in real life). The problem is confounded because the details of how a concept is defined and measured are often not included with a dataset. The problem of semantic interoperability is determining when things are the same, and how much uniformity should be imposed

Semantic interoperability by indicator

One strategy, proposed by the SDG Global Indicator Framework, is to create semantic interoperability at the level of the indicator. The Global Indicator Framework proposes specific measures of items like poverty and food insecurity, that must be measured using the national poverty measure and the Food Insecurity Assessment Survey instrument respectively. Indicator-level semantic interoperability is required to achieve what one respondent described as "comparable fungible metrics at scale across the industry". Interoperability is created when all organizations use a single set of indicators and only these indicators. Annex 1 illustrates examples of semantic interoperability by indicator.

The advantage is that the Global Indicator Framework offers an already-available standard to measure and report and label the goals and targets within the SDGs. The disadvantage of this approach is that the vast amount of already-collected Canadian data pertinent to the SDGs is not compliant with these indicators, so adopting them would mean forfeiting much work to date. Further, and perhaps more problematically, the SDG indicators do not measure the concepts that are meaningful to Canadian organizations in ways that are relevant to Canadians. To adopt these indicators exclusively and in full could create a large volume of perfectly interoperable but mostly useless data.

Semantic interoperability by indicator need not be based on the Global Indicator Framework. Some mission-oriented data ecosystems and many collective impact initiatives begin by convening stakeholders to facilitate agreement on their own bespoke indicators of progress. This helps to pool and prioritize resources; discover and address roadblocks in the system; and leverages data in the context with which it can be most helpful in advancing systemic change. An example is the Youth Investment Fund Learning Project[11] currently underway in the UK. There are also several well-recognized Canadian initiatives.[12] A critical success factor in the development of these ecosystems is to find a balance between the desire to use bespoke indicators, which may be more relevant, with the broader value of standardized indicators. Failing to do so has been shown to result in small pockets of semantic interoperability without addressing the broader potential of national SDG data interoperability.

Semantic interoperability by ontologies

An emerging strategy is to establish semantic interoperability at the level of words, using ontologies. Focusing on ontologies creates the possibility of standardizing the composite elements of indicators, while leaving indicators to be defined by each organization. This is a more flexible approach to semantic interoperability.

The Sustainable Development Goals Interface Ontology (SDGIO) aims to define how "entities relevant to the SDGs can be logically represented, defined, interrelated, and linked to the corresponding terminology in glossaries and resources such as the UN System Data Catalogue and SDG Innovation platform".[13] At present the UN SDG Interface Ontology does not yet provide any semantics necessary for the representation of elements within the indicators, which impedes flexibility, and greatly constrains the power of the Internet of Impact to reason about indicators in meaningful ways.

The Common Approach to impact measurement is a Canadian-led ontology for semantic interoperability of the SDGs. It does provide the semantics necessary to represent the elements within each indicator, allowing for a much more flexible and relevant approach to impact measurement. While still in its early stages, it could be a promising strategy for Canadian SDG initiatives in Canada.

Regardless of the specific standard, ontologies work by reducing the occurrence of both the same things being labelled differently and different things being labelled the same. The ontology does not prescribe a single way of measuring but does make it possible for machines and humans to know when and in what ways things are the same. A good data ontology will allow a human or a machine to recognize different meanings (for example different measures of poverty) without imposing a single measure.

Humans vs Machines

Greater semantic interoperability will allow machines to read and combine datasets without human participation. However, greater semantic interoperability will require a more rigid approach based on a universal set of indicators and measurement methods, which has significant drawbacks. Semantic standards necessarily impose rigidity around what is measured and how, which can limit the scope, range and evolution of data gathered. It can also propagate any biases that are embedded in the measures.

Using semantic interoperability based on ontologies allows greater flexibility. However, this flexibility reduces the degree of interoperability. In the near-term, ontology-based semantic interoperability will not be interoperable enough to leave it to the machines to do the work. Instead humans will be needed to identify when things are the same. If they are the same, machines can easily combine and analyze data. Data ontologies provide humans with the information to assess if differences are material for the purposes that the data is being used for. Where data cannot be made the same, or similar enough to combine, humans can provide nuanced interpretations of the data taking into consideration subtle differences in measurement and definition in different datasets.

Ontology-based interoperability may eventually be sufficient for machines to do the work, but it will take longer than semantic ontology based on indicators. Thus, part of the interoperability question facing Canada is how much flexibility and relevance is the SDG community willing to forgo to quickly usher in an era of machines reading, combining and analyzing SDG data?

Centralized vs Distributed data

A strategic question that Canada's data interoperability strategy faces is how centralized should the data be? The possibilities of joined-up data, and the degree of interoperability needed, varies depending on if Canada strives for data that is distributed, centralized in a single repository, or something in the middle.

Distributed data

Distributed data leaves data within legacy systems. It also allows new data platforms and repositories to arise with little command-and-control coordination. Using a distributed data approach, data continues to be created in a variety of ways and formats. If it is shared, it may be through email or direct link or a link on a repository. Data interoperability can be created by i) applications that can translate between data formats; ii) effective use of metadata standards (e.g. DCAT, CKAN, Schema.org); and iii) some limited semantic vocabularies for enabling semantic mapping done initially by humans (semantic mappings between subsequent versions of the same datasets can be automated using scripts). The advantages are a dynamic and fluid data ecosystem. Data portals can help users find distributed datasets by assembling links and catalogues of data. The disadvantage is that it is labour intensive to keep portals up to date. Today's data ecosystem can be described as distributed data with significant gaps in basic elements of data interoperability, such as limited use of metadata standards and lack of awareness of relevant semantic vocabularies.

With a high degree of flexible semantic interoperability, it will be possible for machines to find, read, and analyze SDG data that is widely distributed on a number of sites. Software applications will aid in the analysis of performance. They will automate the longitudinal analysis (i.e., how and why indicator changes over time) and transversal analysis (i.e., how and why indicators from different locations differ from each other), in order to discover the possible causes of differences. The semantic interoperability enables this to occur, even though the data is widely distributed. However, with no central authority providing the organizational and institutional governance over the data, it will be difficult to enforce a single set of indicators. Thus, for distributed data, an ontology-based approach is likely more feasible.

National data repositories

Some discussions on SDG data imagine a centralized approach: a single SDG data repository either for each nation or globally. A repository is a database that collects data from many sources. In this model, the manager of the repository becomes the data steward in charge of creating and enforcing data interoperability standards, data privacy and security. Organizations submit their data in a manner that complies with rules. The expected advantage of a centralized approach is that data on the repository is easy to find - because it is all in one place - and all data can be compared and compiled with other data on the repository. All data in the repository, for example, could be required to use the Global Indicator Framework. However, if we consider the full range of public and private datasets that might be usefully mobilized toward assessing the progress toward the SDGs, it is difficult to imagine corralling it all into a single national repository (other studies have reached similar conclusions).[14]

Mission Oriented Ecosystems

Mission-oriented data ecosystems are an attractive middle ground. Using this model, communities of stakeholders working toward similar goals create their own data interoperability strategies. With a smaller, more focused community, it is possible to build a high degree of data interoperability. This means that there is tight alignment across all five layers of data interoperability, with an agreed governance model, one or several data repositories with robust data sharing agreements, aligned version control standards, controlled use of vocabularies, and common data formats. The high level of interoperability allows merged data within the ecosystem and has the potential for machine-learning. An example of this is Canada's Homeless Individuals and Families Information System, a data repository and platform that serves an ecosystem of stakeholders around a common set of semantically interoperable indicators. Some are migrating toward ontology-based semantic interoperability to give members of the ecosystem more flexibility to measure what is relevant. Thus, mission-oriented data ecosystems offer an attractive alternative that is neither fully distributed, nor fully centralized, but hubs of centrality.

The Canadian SDG Data Ecosystem

Canada's SDG Data Ecosystem has all the components to make Canada a leader in SDG data interoperability. The Canadian SDG Data Ecosystem, at present, consists mostly of distributed privately-held data, some mission-driven mini-ecosystems, and a few cloud-based measurement platforms that gather SDG data from a wide range of organizations.

The data sources and data providers are organizations that collect data. They include private companies, civil society organizations, universities and individuals. The data may be in standard file formats (for example: an excel file) but may not have adequate metadata, version history and may not be accessible to others. Making this data available will require collaboration with open data organizations, as well as privacy experts and advocates for vulnerable populations to make sure data that should not be made available is not available.

A key driver of data interoperability will be the analysts who are focused on measuring and monitoring progress toward the SDGs. These groups will be key players in linking, visualizing and analyzing data. In the early stages, before semantic interoperability is at a level that allows machines to link datasets, analysts will play an important role using their professional judgement and expertise to know when things are the same.

The data infrastructure will play an emerging role in making data findable and useable and promoting interoperable standards. Data infrastructure refers to portals and repositories, as well as softwares that help organizations measure and track data. Data infrastructure organizations become purveyors of interoperability. They become the organizers of data governance, and they can promote particular technical, syntactic and semantic interoperability.

The mission driven ecosystems within Canada are already creating data interoperability within their communities, however, the resultant pockets of interoperability will not do much to change the overall potential of the coming data revolution. True interoperability will not be created by small communities acting alone. Progress will be made by bringing these communities together around a flexible approach to semantic interoperability.

By good fortune, Canada is home to several of the globally-recognized leaders in semantic interoperability. This presents a unique opportunity for Canada to be a leader in innovative approaches towards data interoperability for SDGs.

Data Sources / Providers

Government/Public

- **StatCan**
- **Government Departments**
- **Health Authorities**

Non-government
Business/Private

Data for SDG Decision making

Business Investment

- Impact investment advisories, e.g. **Rally Assets**
- Responsible investment advisories e.g. **Sustainalytics**
- **Responsible Investment Association (RIA)**
- **Global compact, Canada** #
- Investment Funds and Managers (e.g. **Possibilian Ventures** #)

Government
Non-government
Funders

Analysis, Measurement and Monitoring

Impact Measurement Platforms

- **Impact Dashboard** + #
- **Sametrica** + #
- Dharma.ai
- Vera Solutions Amp Impact #
- Sopact Impact Cloud #
- Sinzer software tool #
- SocialSuite
- Alice.si

Monitoring progress

- **Canada (ESDC, GAC, StatCan, CCIC)**
- **Cities and Communities (CFC, IISD, CCSD, BCCIC)**
- **Business (Global Comp. Canada, Coops and Mutuals)**

Transparency and Accountability

- **Auditor General, provincial and municipal auditors**
- **Accounting firms**
- **Evaluators (Canadian Evaluation Society)**

Data Infrastructure

Portals and Repositories

- **StatCan (SDG Data hub, Census profiles, etc.)**
- QRI
- HIFIS
- Data for all
- Ocean Protocol
- Venture for Good (or source)

Standard Setters

- **Common Approach to Impact Measurement (in development)** #
- ISO
- UNSD #
- **StatCan** #
- Global Compass #
- Impact Measurement Project

Bold denotes a Canadian organization or initiative

+ Denotes alignment with the Common Approach

Denotes SDG-specific features or data

Problems that Data Interoperability does not Solve

In our interviews on interoperable SDG data, many respondents raised general problems with impact data. The scope of this report has been on interoperability and the problems that interoperability can solve. It is important to be clear about the problems interoperability will not solve.

1. Gaps in data collection

Making data more interoperable does not solve for data collection gaps. There are still significant barriers to organizational skills, knowledge and resources (often linked to an absence of incentives) to undertake data collection. With those resources in place, there are challenges gathering data from people that include logistical challenges (keeping track of people, getting them to answer questionnaires) and ethical limits (we need to be cautious about how much time we expect people – especially vulnerable people – to give toward someone else's measurement accountability objectives).

"Genuine engagement with the community is really hard! We want to avoid "mining" people for data."
– Interview Participant

2. Making data more interoperable doesn't render it meaningful or useful

Making data more interoperable doesn't render it meaningful or useful. Interoperable data can give us a much more accurate and consistent picture of "what is," but this cannot tell us "what ought to be." Data interoperability will unlock new analytical insights and enable improved evidence-informed decision making. But data will always need to be interpreted. And the history of science shows us that there are often multiple valid, yet conflicting, interpretations of the same data. Data interoperability can increase what we can do with data, but spaces for deliberation between different interpretations are needed for society to democratically decide what to do about the insights generated from the data.

3. Interoperability doesn't guarantee high quality data

Interoperability doesn't guarantee high quality data. Some of the data standards discussed, such as good providence data and good version control, may equip humans (and perhaps one day machines) to assess the quality of data. For example, some database structures can record information about data validity that remains associated with the data. The would allow the data creator or an analyst to assign a degree of validity or confidence to the data. Further, when data is found to be erroneous, well-linked interoperable data can enable corrections in one dataset to automatically carryforward to other datasets that used the data. However, if the data sources are inaccurate or outdated, joining them up won't solve the underlying data quality issue.

Conclusion: A Leadership Opportunity for Canada

We can leverage Canada's strength in many key areas that would make interoperability for SDG data within reach.

The availability of data around SDGs is not the limiting factor for Canada. Rather, it is how the existing data is integrated and mapped towards a shared, common understanding. Distributed data is an exciting and promising avenue that allows for an ecosystem approach with local relevance. However, even with distributed data, key decisions need to be made at a collective level on the degree of interoperability we seek and on systems of data governance.

Concurrently, advancements can be made at all levels of interoperability. For example, The Common Approach is already progressing on a greater shared understanding of terms and words related to SDGs in the social enterprise sector. The Common Approach[15] is advancing semantic interoperability drawing on an already strong Canadian research community in semantics, machine learning, and impact measurement. Collective impact communities have experience mobilizing mission-driven communities. Groups like FNIGC are innovating world-recognized approaches to data stewardship. Together, these communities can accommodate new approaches to test and scale data interoperability options that allow for distributed data.

Canada is in an ideal position to take global leadership on data for impact through its commitment to advancing the SDGs at home and its internationally recognized capacity in data and information technology. This could be done in an efficient way by establishing an impact data 'taskforce', such as was established for the Pan-Canadian Artificial Intelligence Strategy, perhaps in collaboration with initiatives such as AICommons, and having a mandate for catalysing impact-centred data ecosystems and interoperability. Staffed with a core operating budget and a small team, it could work with stakeholders to pool resources and establish data ecosystems around SDG related priorities for Canada, such as indigenous communities, food security, and energy. In parallel, it could work to pool resources to address cross-cutting issues of interoperability, collaborating internationally to make the deep, long-term investments necessary for robust standards development and adoption.

Interoperability is not a single problem with a single solution. Interoperability requires sustained time, expertise, and material investment to gain and maintain momentum and realize its systemic impact. The nature of interoperability requires all perspectives and full ecosystem engagement: from data generators, users, reporters, consumers and intermediaries, all must be invested in the pathway towards data interoperability. At current, there are many actors and initiatives that are involved in related, but tangential ways, and many are not informed or related to existing initiatives. Those who are aware lack the capacity for alignment to specific systems. Relevancy to the outcomes requires engagement with beneficiaries and practitioners from the beginning. This early-on engagement has the added benefit of limiting and focusing the scope of potential activities and approaches, and re-centers the purpose of enabling better outcomes for people and society. Achieving the future state will be the result of a simultaneous push in multiple areas of data interoperability, and should include the expectation that success will come through trial and iteration, with input from the ecosystem that creates, contributes, and uses data.

"In order to solve our complex challenges we need to move beyond silos of data to connecting it to provide new insights. If Canada gets this right we will enable transformative impact to get us from where we are to where we know we can be."
– Allyson Hewitt, *McConnell Foundation & MaRS Discovery District*

Acknowledgements

Authors

Karen Deng
MaRS Discovery District

Michael Lewkowitz
Possibilian Ventures

Kate Ruff
Carleton University

Alex Ryan
MaRS Discovery District

With contributions from

Mark Fox
University of Toronto

Stefan Jungcurt
International Institute for Sustainable Development

Karl Richter
EngagedX

Laura Schnurr
McConnell Foundation

Mikayla Zolis
MaRS Discovery District

With support from

Employment and Social Development Canada / Emploi et Développement social Canada

La fondation McConnell Foundation

In consultation with

Gina Babinec
Impact Hub Ottawa

Stephen Barnett
Util

Marie-Helene Berburbe
Canadian Audit and Accountability Foundation

Raphael Betti
European Investment Fund

Heather Block
United Way Winnipeg

Greg Bloom
Community Organizer

Tim Dietrich
Kind Village

Donika Dimovska
Earth Security Group

Fiona Duguid
Canadian Co-operative Research Network

Amanda Feldman
Impact Management Project

Delphine Gibassier
Birmingham City University

Sylvia Grewatsch
Ivey Business School

Kasha Huk
BLab Canada

Adam Jog
Mowat Centre

Leslie Jump
Different Funds

Shannon Kindornay
Canadian Council for International Co-operation

Heleigh King
Deloitte

Jean Le Clerc
Kind Village

Michael Lenczner
Powered by Data

Tris Lumley
New Philanthropy Capital

Eleanor Melanson
Statistics Canada

Kate Murray
Rally Assets

Rishbabh Nanda
University of Toronto

Annemarie Naylor
Future Care Capital; Common Futures

Aunnie Patton Power
Intellimpact

Astrid Scholtz
Sphaera

Joelle Sostheim
Pitchbook

References

1. Global Partnership for Sustainable Development Data [Internet]. Interoperability Data Collaborative; 2018 [cited 2019 Apr 29]; [about 1 screen]. Available from: http://www.data4sdgs.org/initiatives/interoperability-data-collaborative
2. Goldstein E, Gasser U, Budish R. Data Commons Version 1.0: A Framework to Build Toward AI for Good. Berkman Klein Center [Internet]. 2018 Jun 21 [cited 2019 Apr 29]. Available from: https://medium.com/berkman-klein-center/data-commons-version-1-0-a-framework-to-build-toward-ai-for-good-73414d7e72be
3. Statistical Data and Metadata eXchange. SDMX 2.1 User Guide [Internet]. SDMX; 2012 Sept 19 [cited 2019 Apr 29]. 145 p. Available from: https://sdmx.org/wp-content/uploads/SDMX_2-1_User_Guide_draft_0-1.pdf
4. Global Partnership for Sustainable Development Data [Internet]. Interoperability Data Collaborative; 2018 [cited 2019 Apr 29]; [about 1 screen]. Available from: http://www.data4sdgs.org/initiatives/interoperability-data-collaborative
5. Data Collaboratives [Internet]. GovLab [cited 2019 April 29]; [about 2 screens]. Available from: http://datacollaboratives.org/
6. United Nations Development Group. Data Privacy, Ethics and Protection [Internet] UNDG; 2017 [cited 2019 Apr 29]. 19 p. Available from: https://undg.org/wp-content/uploads/2017/11/UNDG_BigData_final_web.pdf
7. Tri-Agency Statement of Principles on Digital Data Management [Internet]. Government of Canada; 2016 Dec 21 [cited 2019 Apr 29]; [about 4 screens]. Available from: http://www.science.gc.ca/eic/site/063.nsf/eng/h_83F7624E.html?OpenDocument
8. The First Nations Principles of OCAP [Internet]. First National Information Governance Centre; 2019 [cited 2019 Apr 29]; [about 5 screens]. Available from: https://fnigc.ca/ocapr.html
9. Richter KH. The Internet of Impact [Internet]. UNDP; 2019 Mar 12 [cited 2019 Apr 29]. Presentation: 18 slides. Available from: https://prezi.com/sli1caj-6_ue/internet-of-impact/
10. A New Tool for Data Science [Internet]. Qri [cited 2019 Apr 29]; [about 4 screens]. Available from: https://qri.io/
11. Our Approach [Internet]. Youth Investment Fund; 2016 [cited 2019 Apr 29]; [about 1 screen]. Available from: https://yiflearning.org/our-approach/
12. Cabaj M. Collective Impact & Shared Measurement Tough but Necessary [Internet]. Tamarack Community [cited 2019 Apr 29]. 4 p. Available from: https://www.tamarackcommunity.ca/hubfs/Resources/Publications/Collective%20Impact%20and%20Shared%20Measurement%20Tough%20But%20Necessary.pdf
13. Ontology Lookup Service [Internet]. Sustainable Development Goals Interface Ontology; 2018 Aug 13 [cited 2019 Apr 29]; [about 1 screen]. Available from: https://www.ebi.ac.uk/ols/ontologies/sdgio
14. Scholz A, Selian A, Mohn B. Impact Platforms. [Internet] 2018 Sept 20 [cited 2019 Apr 29]. 55 p. Available from: https://impactfinance.network/wp-content/uploads/2018/10/Impact-Platforms-Report_Final.pdf
15. Common Approach to Impact Measurement [Internet]. Carleton University; 2019 [cited 2019 Apr 29]; [about 1 screen]. Available from: https://carleton.ca/commonapproach/

Annexes

Annex 1: Ontologies for the Representation of UN SDG Indicators

Annex 2: SDG Interoperability Interview Guide

Annex 3: SDG Interoperability Participatory Workshop Methodology

Annex 4: SDG Data Interoperability Literature Summary

Annex 1: Ontologies for the Representation of UN SDG Indicators

This document defines an Ontology for the representation of the UN SDG Indicators (UN SDG, 2018).

Existing Indicator Ontologies

The Sustainable Development Goals Interface Ontology[1] (SDGIO) purports to define "entities relevant to the SDGs can be logically represented, defined, interrelated, and linked to the corresponding terminology in glossaries and resources such as the UN System Data Catalogue and SDG Innovation platform." (UNEP, 2016). A part of the ontology is basically a taxonomy of indicators defined in the UN SDGs. Its relevance lies in that it provides a unique URI for identifying each UN SDG indicator. But it does not provide any semantics necessary for the representation of and reasoning about indicators.

SDMX is syntax for representing and transmitting data. Originally XML-based, and more recently JSON-based. XML is a syntax for representing and transmitting data. Think of it as a grammar. A grammar defines what the structure of sentences are, i.e., noun, verb, etc., and restrictions on the possible orderings they can appear in a sentence. But, a grammar does not define what the nouns, verbs, etc. are. That is, the actual words of the language. For example, British English and Canadian English may share the same grammar, but some of the words are not shared between them, nor are the meanings necessarily shared. For example, the meaning of "public school" in Toronto means a school covering K-6 grades, and the school is funded by the government. In the UK "public school" is a privately funded school for the elite members of society. Same words, different meanings. Another limitation of grammars is that some sentences that are grammatical are semantically inconsistent, e.g., "Colorless green ideas sleep furiously" (due to Chomsky). The purpose of the Ontology is to: 1) identify the concepts of domain (akin to words), and 2) define their meaning. Ontologies address the limitations of syntax above.

IBM's Scribe City Ontology (Uceda-Sosa, et al., 2011) which spans messages, events, city services and city agencies: KM2City, OM/QUDT, ISO, ISO 21972

Proposed Indicator Ontology

In this section we define an Ontology for representing indicator values, meta data, and their definitions. To begin we address the simple requirement of being able to uniquely refer to a specific indicator in the ontology.

Unique Identifiers for UN SDG Indicators

A result of the SDGIO project was the establishment of a unique identifier, i.e., URI, for each SDG indicator. Following are two examples:

SDG ID	URI	Definition
2.1.1	http://purl.unep.org/sdg/SDG-IO_00020014	Prevalence of undernourishment
2.1.2	http://purl.unep.org/sdg/SDG-IO_00020015	Prevalence of moderate or severe food insecurity in the population, based on the Food Insecurity Experience Scale (FIES)
2.4.1	http://purl.unep.org/sdg/SDG-IO_00020020	Proportion of agricultural area under productive and sustainable agriculture

Basic Indicator

The basic Indicator class captures four pieces of information: 1) what is the UN SDG indicator URI being represented, 2) what country is the indicator for[2], 3) the year for which the indicator is measured, and 4) the value of the indicator for the specified country and year.

The basic indicator is quite simple, and does not utilize the power of ontologies and linked data. All properties are data properties[3] except for the indicator ID. This provides for an easy to use representation that data can be mapped into. Existing standards are used where appropriate, such as ISO 3166-1 alpha-3 that provides a 3 letter code for all countries. If we wish to designate provinces, then we can add a property for province using ISO 3166-2.

Class	Property	Value Restriction
Indicator	indicatorID	exactly 1 xsd:string
	organization	exactly 1 xsd:string
	country (ISO3166-1 alpha-3)	exactly 1 xsd:string
	year (yyy-mm-dd)	exactly 1 xsd:string
	value	exactly 1 xsd:string
RatioIndicator	owl:subClassOf	Indicator
	numerator	exactly 1 xsd:string
	denominator	exactly 1 xsd:string

A RatioIndicator is a subclass of Indicator. It is distinguished by having a numerator and denominator for representing more details of the data used to derive the indicator value.

LD Indicator

The LDIndicator extends the Indicator class to include ontology/linked data standards for indicator data. In particular, for each data property in the original Indicator class, there is a corresponding object property[4]. LDIndicator is a subclass of Indicator, which allows for the inheritance of the Indicator data properties into LDIndicator, hence the Indicator data properties co-exist with their corresponding object properties in LDIndicator.

The following table lists the correspondence between data and object properties:

Data Property in Indicator	Corresponding Object Property in LDIndicator	Comment
indicatorID	hasIndicatorID	hasIndicatorID's value is restricted to instances of sdgio:SDGIO_00020015 which is the SDG Indicator class in SDGIO.
organization	forOrganization	forOrganization's value is restricted to instances of organization ontology Organization class
country	forCountry	forCountry's value is restricted to instances of schema.org's Country class
year	forYear	forYear's value is restricted to instances of OWL-Time's DateTimeDescription class
value	hasValue	hasValue's value is restricted to instances of the ISO21972 standard Quantity
numerator	om:numerator	om:numerator's value is restricted to instances of the ISO21972 standard Quantity
denominator	om:denominator	om:denominator's value is restricted to instances of the ISO21972 standard Quantity

The following defines the LDIndicator class with the object properties described above:

Class	Property	Value Reflection
Indicator	owl:subClassOf	Value Restriction
	hasIndicatorID	Indicator
	forOrganization	exactly 1 sdgio:SDGIO_00020015
	forCountry	exactly 1 org:Organization
	forYear	exactly 1 sc:Country
	has Value	exactly 1 ot:DateTimeDescription
LDRatioIndicator	owl:subClassOf	RatioIndicator
	om:numerator	exactly 1 iso21972:Quantity
	om:denominator	exactly 1 iso21972:Quantity

Indicator Meta-Information

Meta-information provides information about the indicator, such as provenance, validity and trust. In the following we extend the LDIndicator to include meta-information as defined in (Fox, 2015). Text in blue are the extensions.

Class	Property	Value Restriction
LD Indicator	owl:subClassOf	Indicator
	owl:subClassOf	iso21972:Indicator
	owl:subClassOf	pr:Entity
	owl:subClassOf	kp:KP_prop
	indicatorID	exactly 1 sdgio:SDGIO_00020015
	forOrganization	exactly 1 org:Organization
	forCountry	exactly 1 sc:Country
	forYear	exactly 1 ot:DateTimeDescription
	hasValue	exactly 1 gci:Quantity

To represent Provenance we incorporate the W3C Provenance Ontology PROV (Belhajjame et al., 2012). The core classes and properties are depicted in Figure 1.

Figure 1: PROV Core Classes and Properties

Simply, activities use entities to generate new entities. Agents are associated with activities and contribute to entities. Details can be found in Belhajjame et al. (2012). These classes and properties enable the specification of the workflow that created the indicator.

An ongoing issue is whether information/data found on a page is correct (true) or incorrect (false). Whether the creator of the information deliberately makes false statements, or unknowingly copies false information from another site, there is no way to discern what is correct from incorrect. The same holds with city indicators. Data and analyses that are believed to be true at the time they are gathered or computed, may be found over time to be incorrect. Or it may not be clear whether the information is true or not, especially if the indicator is based on a sampling of a population, but one can assign a degree of validity to the information. In addition, in the case where data is derived from other data, and the latter is no longer valid at some point of time, then the former becomes invalid for that same point of time.

To represent data Validity we adopt the ontology defined in Fox (2015). By specifying that LDIndicator is a subclass of KP_prop, it inherits the properties of KP_prop as specified below.

Class	Property	Value Restriction
kp:KP_prop	kp:assigned_certainty_degree	exactly 1 xsd:real
	kp:effective	only ot:DateTimeInterval
	kp:is_dependent_on	only kp:KP_prop

The assigned certainty degree specifies a number between 0 and 1 as being the probability that the indicator value is correct. The effective property specifies the time period during which the certainty degree is valid. The property is dependent on specifies the other KP props that the certainty degree of the indicator is dependent on.

Indicator Definitions

As the data used to derive indicators is made available on open data sites, it enables the development of software applications that will aid in the analysis of performance. In particular, it becomes possible to automate the longitudinal analysis (i.e., how and why indicator changes over time) and transversal analysis (i.e., how and why indicators from different locations differ from each other), in order to discover the possible causes of differences.

But the assumption that organizations will adhere to the standard is a strong one, as organizations often interpret definitions differently. Before any meaningful analysis can be performed, three questions with respect to consistency need to be answered: Is an organization's interpretation of an indicator:

1. Definition consistent, e.g., is the definition of an indicator reported by an organization consistent with the indicator's definitions?
2. Intra-indicator consistent, e.g., is the data used to derive an indicator from the same time and location?
3. Inter-indicator consistent, e.g., are the definitions of the things being measured consistent across time?

Up to this point, the Indicator and LDIndicator classes have focused on representing the value of an indicator, and any meta-information (i.e., provenance and validity) available. ISO 21972 provides an ontology that can be used to represent the definition of an Indicator. The GCI Foundation Ontology provides the details upon which ISO 21972 is based (Fox, 2013; 2015). To allow for the definition of an indicator, LDIndicator is extended by making it a subclass of iso21972:Quantity.

Class	Property	Value Reflection
LDRatioIndicator	owl:subClassOf	Indicator
	owl:subClassOf	iso21972:Indicator
	owl:subClassOf	pr:Entity
	owl:subClassOf	kp:KP_prop
	indicatorID	exactly 1 sdgio:SDGIO_00020015
	forOrganization	exactly 1 org:Organization
	forCountry	exactly 1 sc:Country
	forYear	exactly 1 ot:DateTimeDescription
	hasValue	exactly 1 gci:Quantity

Example

In this section we show how the indicator 2.4.1, "Proportion of agricultural area under productive and sustainable agriculture", is represented. Let's assume we have an indicator for an organization called "Sustainable Farming Coop" who has 100 acres of land, of which 25 acres are productive and sustainable.

The basic Indicator representation without using linked data is:

Instance	Property	Value
SFC_2.4.1	rdf:type	Indicator
	indicatorID	"2.4.1"
	organization	"Sustainable Farming Coop"
	country (ISO 3166-1 alpha-3)	"Canada"
	year (yyyy-mm-dd)	"2019-03-26"
	value	0.25

Since the 2.4.1 is a ratio, we could have used the RatioIndicator class to include the numerator and denominator:

Instance	Property	Value
SFC_2.4.1	rdf:type	RatioIndicator
	indicatorID	"2.4.1"
	country (ISO 3166-1 alpha-3)	"CAN"
	year (yyyy-mm-dd)	"2019-03-26"
	value	0.25
	numerator	"25"
	denominator	"100"

While this is the simplest representation of an indicator's value, it leaves a lot of information out, such as the units of the numerator and denominator (i.e., acres), and any meta information. The first step to representing this information is to substitute the Indicator class with the LDIndicator class, and convert the values into instances. Let's assume the following:

- The URI for indicator 2.4.1 is http://purl.unep.org/sdg/SDGIO_00020020
- The URI for "Sustainable Farming Coop" is http://SustainableFarmingCoop.com.
- The URI for Canada is gn:6251999
- The URI for the year 2019 is dt2019
- The URI for the value is val_sfc_2.4.1_2019

Class	Property	Value Reflection
SFC_2.4.1	rdf:type	Indicator
	rdf:type	iso21972:Indicator
	rdf:type	pr:Entity
	rdf:type	kp:KP_prop
	indicatorID	http://purl.unep.org/sdg/SDG-IO_00020020
	forOrganization	http://SustainableFarmingCoop.com
	forCountry	gn:6251999
	forYear	dt2019
	hasValue	exactly 1 iso21972:Quantity
dt2019	rdf:type	ot:DateTimeDescription
	year	2019
quant_sfc_2.4.1_2019	rdf:type	iso21972:Quantity
	iso21972:value	val_sfc_2.4.1_2019
val_sfc_2.4.1_2019	rdf:type	iso21972:Measure
	iso21972:numerical_value	0.25
	iso21972:unit_of_measure	xsd:real

If we wish to add provenance and validity meta-information, SFC_2.4.1 would be extended as follows:

Class	Property	Value Restriction
SFC_2.4.1	rdf:type	Indicator
	rdf:type	iso21972:Indicator
	rdf:type	pr:Entity
	rdf:type	kp:KP_prop
	indicatorID	http://purl.unep.org/sdg/SDG-IO_00020020
	forOrganization	http://SustainableFarmingCoop.com
	forCountry	gn:6251999
	forYear	dt2019
	hasValue	exactly 1 iso21972:Quantity
	kp:assigned_certainty_degree	0.75
	pr:wasContributedBy	http://SustainableFarmingCoop.com

Prefixes

Prefix	URI
gn	http://www.geonames.org/ontology/ontology_v3.1.rdf#
gci	http://ontology.eil.utoronto.ca/GCI/Foundation/GCI-Foundation.owl#
iso21972	TBD
kp	http://ontology.eil.utoronto.ca/kp#
org	http://ontology.eil.utoronto.ca/organization.owl#
om	http://www.wurvoc.org/vocabularies/om-1.8/
ot	https://www.w3.org/TR/owl-time/
pr	http://www.w3.org/ns/prov#
sc	http://schema.org/
sdgio	http://purl.unep.org/sdg/
tr	http://ontology.eil.utoronto.ca/trust.owl#

Footnotes

[1] https://www.ebi.ac.uk/ols/ontologies/sdgio
[2] The class of URI for a country has not been selected yet as the FAO ontology (UN Food and Agricultural Organization), which provides URIs for countries, is not available online (http://www.fao.org/countryprofiles/geoinfo/en/).
[3] A data property is one whose value is a literal, namely a number, string (alphanumeric), or date/time.
[4] An object property's value is another object, which is the URI for another class or and instance of a class.

References

Belhajjame, K., Deus, H., Garijo, D., Klyne, G., Missier, P., Soiland-Reyes, S., and Zednik,S., (2012), "PROV Model Primer", http://www.w3.org/TR/prov-primer.

Fox, M.S., (2013), "A Foundation Ontology for Global City Indicators", Global Cities Institute", University of Toronto, Working Paper 3. Downloaded from: http://eil.mie.utoronto.ca/wp-content/uploads/2015/06/GCI-Ontology-v10.pdf

Fox, M.S., (2015), "The Role of Ontologies in Publishing and Analyzing City Indicators", Computers, Environment and Urban Systems, Vol. 54, pp. 266-279.

SDMX, (2012), "SDMX 2.1 User Guide". Downloaded from: https://sdmx.org/wp-content/uploads/SDMX_2-1_User_Guide_draft_0-1.pdf

Uceda-Sosa, R., Srivastava, B., and Schloss R.J., (2011), "Building a Highly Consumable Semantic Model for Smarter Cities", Proceedings of the Conference on Artificial Intelligence for a Smarter Planet, AAAI, p. 1-8.

UN SDG, (2019), "Global indicator framework for the Sustainable Development Goals and targets of the 2030 Agenda for Sustainable Development", UN Document A/RES/71/313 E/CN.3/2018/2. Downloaded from: https://unstats.un.org/sdgs/indicators/Global%20Indicator%20Framework%20after%20refinement_Eng.pdf

UNEP, (2016), "Clarifying terms in the SDGs: representing the meaning behind the terminology", Downloaded from: https://unstats.un.org/sdgs/files/meetings/iaeg-sdgs-meeting-02/Statements/UNEP%20-%20Clarifying%20terms%20in%20the%20SDGs.pdf

Annex 2: SDG Interoperability Interview Guide

Strategy

Individual interviews and community workshops to probe for diverse perspectives and insights. Semi-structured interviews with practitioners and ecosystem participants

- Phase 1: Interviews to uncover core themes and identify other key participants
- Phase 2: Follow-up with key participants identified in Phase 1 as appropriate
- Phase 3: Engage key stakeholders to review proposed direction and begin to identify follow-on opportunities and collaboration

Approach

The goal of these interviews is to determine the barriers to achieving impact data interoperability and it's broad based, high value usage. Participants will be oriented to different positions on the data generation and usage value chain but may have relevant insights regarding multiple aspects. As such, it's recommended that, regardless of their position, we probe their perspectives across the value chain.

Value chain:

- Data origination and collection
- Data formats and standards
- Data sources and intermediaries
- Data combination and analyses
- Use cases
- Outcomes
- Impacts

Sample question guide

- How does your work relate to impact data?
- How is impact data used as a result of your work?
- What sort of decisions and outcomes happen as a result of that usage?
- Where does the impact data you use come from?
- What standards, sources, and intermediaries are involved?
- What are most notable limitations in the impact data you use and have access to?
- What would you be able to do if those limitations were removed?
- What is it that leads to those limitations?
- What would most help resolve them?
- If those things were solved, what's your vision for how impact data could be used and shape the world in the next 5-10 years?

In the responses to the questions, look to also tease out implications across these categories:

- Technical
- Data and format
- Human
- Institutional and organizational

Technical and Setup Notes for Interviewer

- Where possible, use a service to generate and audio file that can be transcribed (e.g. zoom + sonix for transcription)
- Ask participant for permission to record
- Ask participant permission to be attributed in quotes

Annex 3: SDG Interoperability Participatory Workshops Methodology

Strategy

Conduct three structured participatory online workshops with practitioners and ecosystem participants categorized thematically by use cases.

1. Policy & Planning
2. Operations & Audit
3. Funding, Action & Engagement

Participatory: Engage participants by structuring the online workshop around questions, encouraging participants to log their answers through the Poll Everywhere software. Categorize participants by roles and organizations.
Online: Engage a broad geographical range of national and international participants

Participant Categories

Webinar 1: Policy & Planning
When: Monday April 15th, 10am - 11am Eastern Daylight Time; 2pm - 3pm (UTC)

Webinar 2: Operations & Audit
When: Tuesday April 16th, 1pm - 2pm Eastern Daylight Time; 5pm - 6pm (UTC)

Webinar 3: Funding, Action & Engagement
When: Wednesday April 17th, 2pm - 3pm Eastern Daylight Time; 6pm - 7pm (UTC)

Approach

The goal of these online workshops is to identify beneficiaries top impact data sharing needs. We want the participants to feel heard and that their time and perspective was respected by contributing their local needs and barriers from their lived experience.

- Conduct over one hour
- Start with welcome, context and overview of the project
- Ask the following questions through the Poll Everywhere software
  - What are the top 10 questions you ask yourself every day that you can't get an answer to?
  - What type of data do you think you need?
    - Prompt sources, sets
    - Prompt time frequency
    - Prompt geographic scope
    - Prompt accuracy, latency, fidelity
  - What stops you from getting that data?
  - If data was more [shareable, comparable, verifiable, aggregatable] what could you do differently?
  - What data do you have that you think could be of value to others?

Technical and Setup Notes for Interviewer

- Expected Number of Participants per Webinar: 5-15
- T- 60m: Test technology and Breakout room functionality
- T- 30m: Facilitation team connects for final runthrough
- T- 10m: Open line and press RECORD

Annex 4: SDG Data Interoperability Literature Summary

Data interoperability can be defined as "the ability to join-up and merge data without losing meaning". It allows different information systems to work together, which is especially important in the development context, where data is shared internationally but largely collected domestically. Goldstein's Data Commons Framework divides interoperability into a series of layers that can be simply summarized as the technology layer, the data and format layer, the human layer, and the institutional and organizational layer. The technology layer is the most basic and outlines the fundamental concept that data must be published and made accessible on the Internet. The data and format layer focuses on the need for consistency in data structures, standards, and vocabularies. Similarly, the human layer is based on the common understanding of individuals who produce and use the data. Finally, the institutional and organizational layer covers the highest level of data interoperability, including the responsibility to collect data, accountability for accurate and consistent data, and high-level policies such as data sharing agreements. Together, these layers show the elements that are needed for successful data interoperability.

Technology

The literature around technology covers a broad range of technical aspects surrounding data interoperability. Simple but recurring concepts in the research include the need for data to be presented in a machine-readable way, and to include metadata for the important context it provides. In one study about existing models of interoperability in healthcare data, there is some discussion around the technical infrastructure needed to achieve these models. The point-to-point oriented model allows independent development of an information system, aligning terminology, messaging protocol and business processes. The standard-oriented model requires a common standard to be followed, but it is difficult to reach agreement, from the human and institutional perspectives, on those standards. The common-gateway model converts inconsistent data inputs into a standard format and allows communication between information systems. A separate model, the Generic Statistical Business Process Model, defines the business processes needed to produce official statistics, providing a standard framework and terminology. It can also be used as a template for integrating data and metadata standards and harmonising statistical computing infrastructures.

New data sources and technologies will be important to achieving interoperability, according to the Dubai Declaration. Similarly, the Cape Town Global Action Plan highlights in its second strategic area the need for innovation and modernization of national statistical systems. Other strategic areas in the action plan address statistical capacity building, which refer to both technical and human aspects. Within strategic area 2, the action plan proposes to facilitate new technologies and data sources into mainstream statistical activities. To achieve this, the plan includes identifying specifications for flexible information systems needed to allow the strategic use of new and emerging technologies; identifying and removing barriers to use of new data sources while coordinating efforts to incorporate them into mainstream programmes; developing guidelines on the use of new and innovative data generated outside the official statistical system into official statistics; and promoting the development of integrated database systems to support the review and follow up of the implementation process of the SDGs. A UN stats report identified standardised interfaces as one of the key dimensions to enable data interoperability, among other, mostly format related, dimensions.

An existing technology facilitating similar work is the Simple Knowledge Organisation System, used to define relationship between concepts within and across concept schemes, working on the Joined-Up Data Standards (JUDS). It also hosts a project by the Natural Resource Governance Institute, containing government identifiers. This technology exports formats that are conducive to data visualizations like Chord and Sankey diagrams, to simplify the information. The Data For All toolkit (DFA) is a collaborative platform of interoperable software tools, intended to help produce reliable and timely data. It is already being used to help member states strengthen data and statistical systems to inform evidence-based policy decisions that will contribute to the SDGs. While new technologies are often cited as enablers of data interoperability in the literature, inflexible ageing technology environments are identified as a threat to statistical organizations, and challenges associated with harnessing alternative data sources like sensors and satellites.

At the intersection between the technology and data layers, syntaxes and vocabularies contribute to both the technological systems and the standard formatting of data. Information made available on XML identifies it as a technical syntax used to format data in a consistent way, but there are limitations around the syntax as it identifies the concepts of a domain without defining their meaning. This limitation can be compared to using consistent grammar and sentence structures without a common definition for the words being used. This source proposes developing an ontology that would identify the domain and define their meaning, rather than relying on a syntax. DCAT is another vocabulary, grounded in Dublin Core, that is considered flexible and practical when dealing with metadata. The standard is good for use with datasets published by a single source, but it is more complicated when applied to larger data warehouses. Resource Description Framework (RDF) Data Cube Vocabulary is another core metadata standard. Platforms used to publish open metadata are most often Socrata, CKAN, DKAN, Junar, and ArcGIS open data.

Data & Format

Literature relating to data and format is consistent in citing the need for standards and in highlighting the barriers to creating universal standards. The use of common classifications is important to ensure that similar and comparable information isn't classified differently and missed or duplicated in machine-readable data. Conversely, trying to apply a standard in too many contexts creates complex answers to specific simple needs. The use and re-use of existing standards, rather than the creation of new standards, is ideal as an essential step to implementing interoperability. Certain requirements should be met before creating a new data standard, including: a clear need and demand; avoiding duplication or competition with other standards; intellectually, logically, and methodologically sound design; components within the standard adopt existing standards; comparable and interoperable with other standards; data availability through open, sustainable, and easily accessible channels; political buy-in from institutions that will drive the standard and produce the data; realistic timelines for development, implementation and adoption; existing and historical data can feed the standard. Many of these are reflected in the questions used in the JUDS consultation, with some additions around the integration of international standards, the standard-setting process, global strategies on harmonisation and knowledge sharing, as well as sector specific challenges, and effective technical approaches. These are largely theoretical ideals, but in practice, standards change and are often observed for only a short time. When they are discarded for an alternative standard, they do not cease to exist, but instead continue to crowd the space of data standards.

In order to ensure datasets can be integrated with one another across information systems while retaining their meaning, multiple stakeholders must agree to follow common principles and procedures that allow for data standardisation, comparability and integration. The Dubai Declaration includes standards that stress quality, timely, relevant, open and disaggregated data. The importance of disaggregated data is echoed in several sources as essential to ensure that no one is left behind. The activities laid out in the Cape Town plan include modernizing statistical standards, particularly those facilitating data integration and automation of data exchange across different stages of statistical production, as well as defining and implementing standardized structures for the exchange and integration of data and metadata. A UN stats report identifies 5 dimensions that will allow interoperability of data, including one around canonical data and metadata models, another on classifications and vocabularies, and a third on implanting linked-data approaches.

Human

The human element of data operability focuses on the users and producers of data and standards. Roger's diffusion of innovation model can be applied to the considerations that should go into creating a standard, including relative advantage, compatibility, low-complexity, trialability, and delivering benefits that are observable. Both a human and institutional issue, the political support must be fostered for interoperability issues, but this presents a particular challenge of persuading people of the benefits without having any evidence of the future value. The literature around this value claims that interoperability solutions can simplify the process for users to access, share, manipulate, and use complex data, helping to achieve and monitor the SDGs. Similarly, the Dubai Declaration acknowledges the need for trust among data producers and users in order for statistical systems to function at the level of the UN Fundamental Principles of Official Statistics. The Cape Town action plan includes a human aspect in strategic area 4, discussing the dissemination and use of sustainable development data. While human elements are some of the key enablers of data interoperability, they are also key barriers, with the rigidity of established processes and methods acting as an immediate threat, and a lack of skilled workers as an emerging one.

Institutional & Organizational

Committee standards are determined through consensus by a consortium, but they do not usually become the presumed tool. This process, while democratic, risks favouring particular companies, technologies or markets, unless the group is distant from implementors, which instead increases the consortium's struggle for adoption or diffusion of new standards. While standards should be set internationally for consistency, data collection must be driven by subnational needs, and political support should extend to the empowerment of governments to use the data at national and subnational levels. A UN stats publication encourages the investment of time and resources in developing and deploying an interoperability solution, flagging it as an opportunity to make better use of data that already exists in sectoral or institutional silos. Another UN stats report, which outlines dimensions that will enable data interoperability, includes an area on data management, governance and interoperability. Governance and leadership come up repeatedly in the literature as a requirement for successfully joining-up data and enabling interoperability.

The political support mentioned above needs to include efforts to collaborate and share knowledge without duplicating work, as well as encouraging synergies across communities, specifically between technical and policy stakeholders. The Dubai Declaration mirrors the sentiment of cooperation, predicting that the data demands of the 2030 agenda will require partnerships between national statistical authorities, the private sector, civil society, and academia. Cape Town's action plan also includes partnerships in strategic areas 1 and 5, focusing on 'coordination and strategic leadership on data for sustainable development' and 'multi-stakeholder partnerships for sustainable development data'. Strategic areas 3 and 6 address institutional level capacity building, including the strengthening of statistical activities and programmes, with particular focus on monitoring for the SDGS, as well as mobilizing resources and coordinating efforts for statistical capacity building. An activity under the second strategic area of the Cape Town plan aims to modernize governance and institutional frameworks to allow national statistical systems to meet the demands and opportunities of the evolving data ecosystem, as well as defining and supporting the role of the national statistical systems in open data initiatives. A report from JUDS identifies five frontiers for data interoperability policy, including the consolidation of existing principles, shifting from problem identification towards solutions through increasing numbers of data initiatives, turning big ideas into practical solutions, producing global standards and regulations that work for everyone, and creating new adaptable partnerships.

The idea of interoperability has been applied beyond data, like the market network design of Project 1800 that aims to standardize the impact investing process and become compatible with other funding and investment tools. Holders of private data have also made efforts towards joining-up data and promoting interoperability. The Hewlett Foundation and the GovLab have launched a project called DataStewards.net to identify and connect existing practices of data stewardship and to define new approaches to data responsibility. Data collaboratives are public-private partnerships that allow for collaboration across data sources, sectors, SDGs and geographies, but these are still rare and small, largely because corporations lack a mandate to harness the potential of their data towards positive public outcomes.

Further Readings

Project 1800
http://www.global-geneva.com/project-1800-saving-the-sdgs-and-the-world/

UNSC: Where Next for SDG Data Interoperability?
https://unstats.un.org/unsd/statcom/50th-session/side-events/20190304-1M-Where-next-for-sdg-data-interoperability.pdf

Dubai Declaration
https://undataforum.org/WorldDataForum/wp-content/uploads/2018/10/Dubai_Declaration_on_CTGAP_24_cto-ber-2018_online.pdf

Cape Town Action Plan for Sustainable Development Data
https://unstats.un.org/sdgs/hlg/cape-town-global-action-plan/

Inclusive Data Charter: Data4SDGs
http://www.data4sdgs.org/initiatives/inclusive-data-charter

Principles to Practice : a consulation on joined-up data standards
http://www.data4sdgs.org/initiatives/inclusive-data-charter

Development Initiatives: Joined up data standards
http://devinit.org/post

The Problem Statement
https://statswiki.unece.org/display/CSPA/The+Problem+-Statement

Data for All Toolkit
http://dataforall.org/

For full anotated bibliography, please contact the authors of this report