National Standard of Canada
Standards Proposal

Proposed Standard Title:
Series of standards for the sharing of urban dataset meta-data

Plain Language Summary of Standards Proposal (200 words max):
Currently urban data is created and generated from wide range of sources including physical, digital, environmental, social and civic. This has led to the creation of open and closed dataset repositories which are used for a variety of government, commercial, social and civic purposes. A key challenge is to know what urban data exits and finding where the relevant datasets reside in disparate data repositories and systems. The proposed series of Standards will enable greater awareness and discoverability of relevant datasets by enabling organizations to communicate information about the datasets they possess. This will enable a network of linked dataset catalogues across Canada for those searching for relevant urban data. The standard will also specify an adoption maturity model that provides dataset providers guidance on the information to provide about their datasets at each level of maturity.

Proposed Scope:
The proposed series of Standards will specify minimum requirements to enable sharing of dataset catalogue meta-data across multiple data management platforms. Considerations will include: 1. Basic information about the dataset, including publisher, publication date, and domain. 2. Information about the data model used by each dataset, including definitions of entities and attributes. 3. Information about the organization that produced the dataset. 4. Information about the provenance of the dataset. 5. Information about the quality of the dataset. 6. Information about who, how, when and for what purpose the dataset has been used. **Part 1 of the series -** will build on the W3C Data Catalog Vocabulary (DCAT version 3)[1] by integrating and extending other relevant standards such as: • DDI[2]: Data Documentation Initiative standard for documenting statistical data sets.

---

[1] https://www.w3.org/TR/vocab-dcat-3/

[2] https://ddialliance.org/

- PROV-O[3]: W3C standard for representing data provenance.
- The Organization Ontology[4]: W3C standard for representing Organization structure.
- DQV[5]: W3C standard for data quality.
- ISO/IEC AWI 5087-2 standard (under development) for representing City data.

Part 2 of the series - will be an extension of Part 1 and will incorporate OCAP[6] principles in the meta-descriptions of indigenous data.

Part 3 of the series - will be an **adoption maturity model** that identifies the minimum types of meta-data for each level of maturity, to offer guidance as to which portions to use in defining a dataset's meta-data.

Strategic Need:
Identify the strategic need of key stakeholders and confirmation expressing the need.

This includes consideration for:
a. *The strategic need of key stakeholder (e.g. legislator, industry, government, consumers);*
b. *The type of standard (international, regional, domestic standards and harmonization need);*
c. *Addressing up-to-date vs outdated standard to ensure latest innovative/technology/safety features available for businesses;*
d. *If the standard is intended to support national/regional/international certification programs;*
e. *If there is stakeholder intention to transition to different standard;*
f. *The type of maintenance (periodic, continuous, stabilized, best before date); and*
g. *The use of "CAN" descriptor.*

Currently urban data is created and generated from wide range of sources including physical, digital, environmental, social, and civic. This has led to the creation of disparate urban datasets that are not linked in any way. This has resulted in an 'awareness problem as vast majority of these datasets "lie fallow" as researchers and practitioners are unaware of its existence let alone its possible relevance. The challenge is how to enable dataset consumers to become aware of the relevant datasets created by dataset producers. Conversely, how can dataset producers make their existence known?

To address the urban data awareness problem, many government and non-government organizations have or are working towards the creation of city data repositories. Many cities offer open data portals that make municipal data available for a variety of commercial, social,

---

[3] https://www.w3.org/TR/prov-o/

[4] https://www.w3.org/TR/vocab-org/

[5] http://www.w3.org/ns/dqv

[6] First Nations' principles regarding the collection use and disclosure of data or information regarding First Nations.

civic and academic purposes. Some countries have national repositories of urban data, such as the UK's Urban Big Data Centre[7], and Australia's Urban Research Infrastructure Network[8].

The proposed series of Standards will address the following strategic needs:

1. **Data Consumers** will have the ability to search and discover relevant urban datasets for a variety of academic, government, commercial, social and civic purposes
2. **Data Producers** will have the capability to communicate the existence and relevance of their content dataset
3. **Data Aggregators** will have the ability to foster a network of dataset catalogues and will serve as a hub for dataset consumers and producers.

The proposed series of Standards will also foster a set of online tools that dataset creators, whether they be in academe, industry, NGOs or government, will use to document and communicate their datasets.

Adoption of the proposed series of Standards will advance Canada's leadership role in the creation of city-wide common data models whose goal is to break down the traditional 'silos' of city data incompatibility by facilitating the sharing and exchange of city data within the city, between cities and external systems. A city-wide common data model is a core foundation the design, planning and operations of responsible smart cities whose goal is to improve the delivery and impact of programs and services to its citizens.

The need for such Standards as a basis for solving the awareness problem has been verified by a series of panel discussions[9] with Subject Matter Experts, who are both producers and consumers of datasets.

Need for Availability in Both of Canada's Official Languages:	Y
Do stakeholders need the standard published in both official languages?	
Do users of the standard need the standard published in both official languages?	
Do authorities having jurisdiction need the standard published in both official languages?	
Are there health and safety related needs for the standard to be published in both official languages?	
For adoptions, is there availability of the regional/international standard or other deliverable from the source?	
For adoptions, is there an agreement with the source committee to facilitate official translation?	

---

[7] https://ubdc.ac.uk/

[8] https://aurin.org.au/

[9] Undertaken by the Urban Data Centre at the University of Toronto's School of Cities.

Geographical Representation Considerations:
Identify the Canadian geographical representation appropriate to the subject area covered by the standard.

Geographic representation may consider factors such as:
   a. Industry (e.g. petroleum in petroleum producing provinces);
   b. Reference in regulation (if a regulation exists in a province); or
   c. Commodity characteristics and social impact (e.g. heating oil for northern climates).

The proposed standard spans all sectors/domains and is Canada wide.

Trade:
Identify how the standard meets the needs of the marketplace and contributes to advancing trade in the broadest possible geographical and economic contexts.

For example:
   a. Facilitate Canadian innovation to lead internationally;
   b. Support the objectives of "One standard, one test, accepted everywhere";
   c. Support the objectives of "First to Market"; or
   d. Foster international/ regional/ national alignment of requirements.

Urban data is key to supporting the evolution of prosperous, sustainable and inclusive smart cities through technology, government, social and civic innovation. For example, according to Richard Florida, Professor at the Rotman School of Management "urban-tech investment totaled more than $75 billion over this three-year period, representing roughly 17 percent of all global venture-capital investment. Between 2016 and 2017, urban-tech investment more than doubled—from less than $20 billion to $44 billion—as its share of global venture investment surged from 13 percent to 22 percent. Urban tech may well be the largest sector for venture capital investment, attracting considerably more funding than pharma and biotech ($16 billion in 2017) or artificial intelligence ($12 billion in 2017).

Relevant existing documents at the international, regional and national level:
   • W3C Data Catalog Vocabulary (DCAT version 3)[10]
   • DDI[11]: Data Documentation Initiative standard for documenting statistical data sets.
   • PROV-O[12]: W3C standard for representing data provenance.
   • The Organization Ontology[13]: W3C standard for representing Organization structure.
   • DQV[14]: W3C standard for data quality.

---

[10] https://www.w3.org/TR/vocab-dcat-3/

[11] https://ddialliance.org/

[12] https://www.w3.org/TR/prov-o/

[13] https://www.w3.org/TR/vocab-org/

[14] http://www.w3.org/ns/dqv

- ISO/IEC AWI 5087-2 standard (under development) for representing City data.

Institutional and Organizational Support for Standard(s)

Below is a list of institutions/organizations that have expressed support for the Standard
- Innovate Cities
- Esri Canada
- Public Health Agency Canada
- Toronto Public Libraries
- University of Toronto Libraires
- Federation of Canadian Municipalities
- Statistics Canada (Pending)
- First Nations Information Governance Council (Pending)