Towards an International Standard for Transportation Planning Data

Webinar Series: June 5, 2020

Canadian Transportation Research Forum Annual Conference

Megan Katsumi (katsumi@mie.utoronto.ca)

Mark Fox (msf@mie.utoronto.ca)

UNIVERSITY OF TORONTO
FACULTY OF APPLIED SCIENCE & ENGINEERING
Transportation Research Institute

Acknowledgements

- This project was supported by the Ontario Ministry of Research and Innovation through the ORF-RE program.

Background: iCity TPSO

- iCity Transportation Planning Suite of Ontologies
- Developed for transportation planning activities in iCity-ORF project
- Led to the creation ISO standards projects for a city data model
  - Initial focus: transportation planning

Wait – what's an ontology?

- A specialized model
  - What are the core concepts and properties that span the domain's data?
    - To what extent can we generalize them in a useful way?
  - What are the key distinctions?
    - Can we formally define necessary and/or sufficient conditions (using properties) for something to be an example (member) of a concept?
- Provides a precise, formal representation that is machine-interpretable
  - More than a reference model (vocabulary) for a domain
  - Supports:
    - Data reuse
    - Data validation
    - Semantic integration
    - Inference

Example: Definition of a Route

- **Simulation Route**: a route represents a possible path of travel. It begins and ends at some distinct nodes in the transportation network and connects the start and end points by accessing some set of arcs.

$$SimulationRoute \sqsubseteq \exists hasStart.Node \sqcap \exists hasEnd.Node \sqcap \exists accesses.Arc$$

- **Wayfinding Service Route**: a route represents a possible path of travel. It has a start and an end node and some associated geometry.

$$WayfindingRoute \sqsubseteq \exists hasStart.Node \sqcap \exists hasEnd.Node \sqcap \exists hasShape.Geometry$$

Data to Knowledge

Semantic Mapping to Ontology

Travel simulation routes

ID	Start	End	Arc	Attr
route1234	N501	N320	A7952	…
…				

{ID} → Simulation_Route

{Start} → Node

Simulation_Route hasStart {Start}

…

Wayfinding Route Records

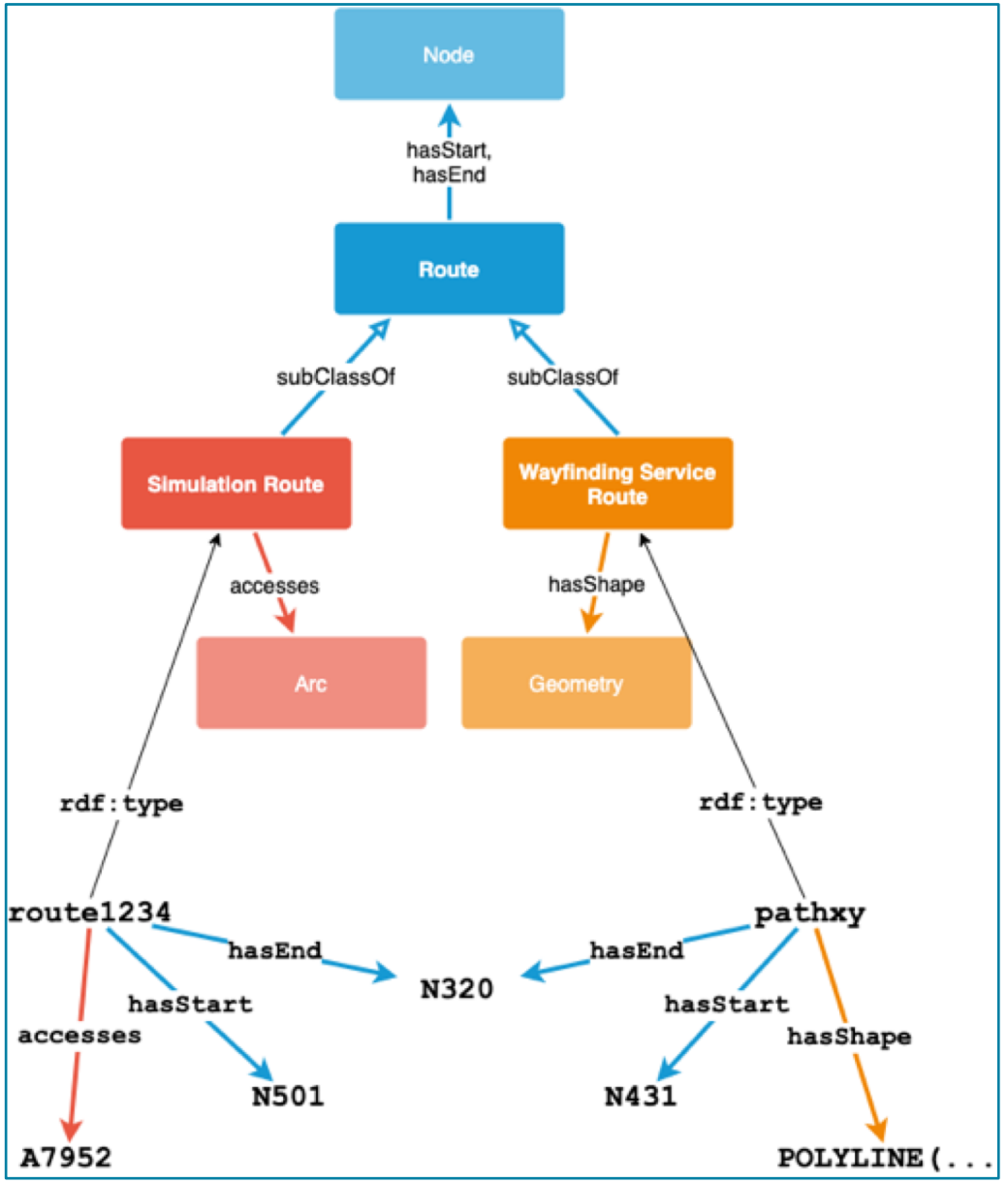
User	Route	o	d	Shape
jsmith	pathxy	N431	N320	POLYLINE(…
…				

Route → Wayfinding_Service_Route

o → Node

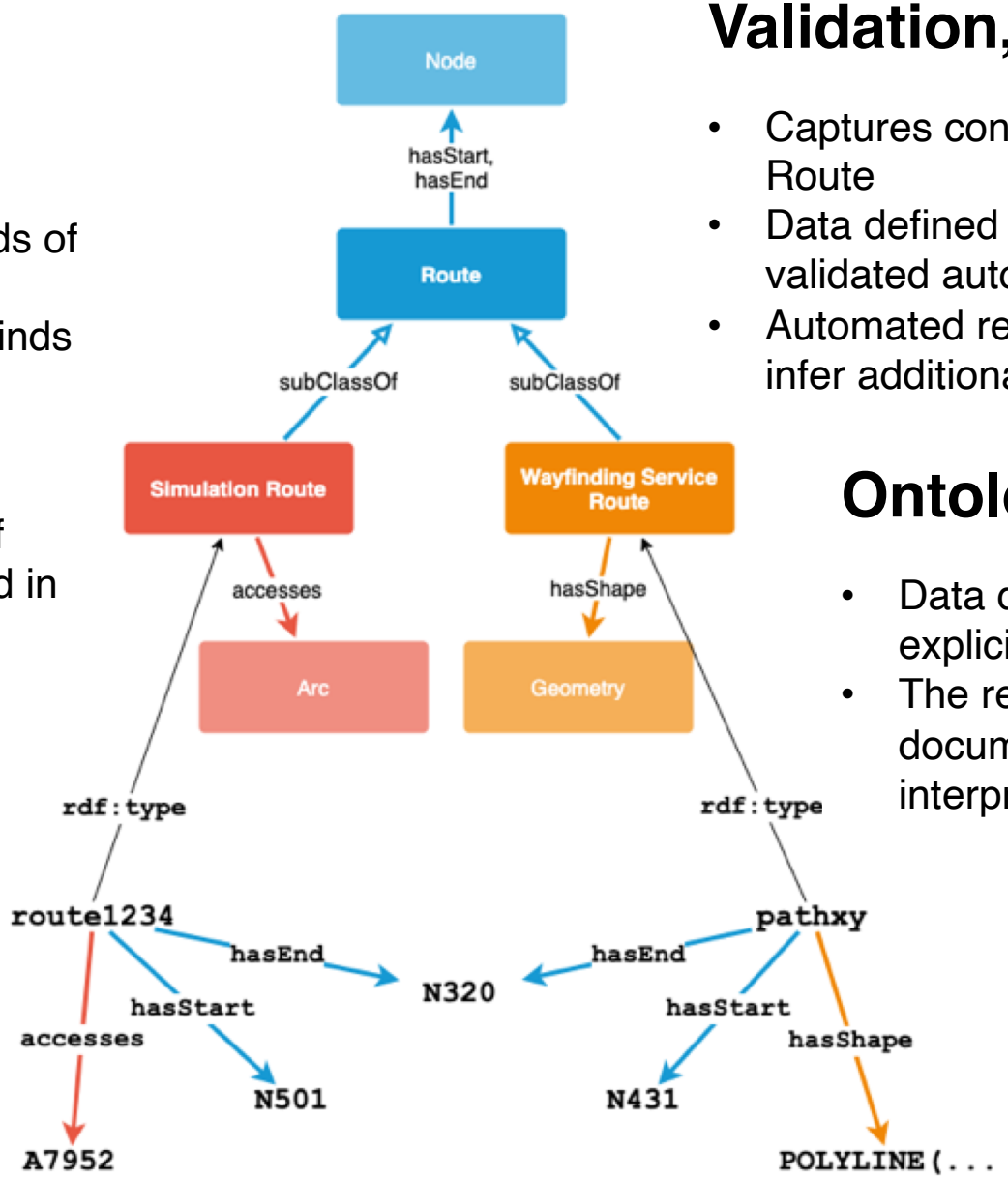
…

Knowledge Graph



Ontology for Integration

- Supports the definition of multiple kinds of route
- Explicitly identifies how the different kinds of routes are related
  - What's common between them
  - What's different between them
- Data sources using either definition of route can be understood and captured in an integrated knowledge base



Ontology for Data Validation, Inference

- Captures constraints on each type of Route
- Data defined with the ontology can be validated automatically
- Automated reasoners can be applied to infer additional knowledge about the data

Ontology for Data Reuse

- Data defined with the ontology has an explicit semantics
- The representation serves as documentation for how it should be interpreted

Our claim

Ontologies (the iCity TPSO in particular) provide a way to address:

- a major challenge for transportation planning,
  and
- a limitation of traditional approaches to standards specification

Transportation Planning: The Data Problem

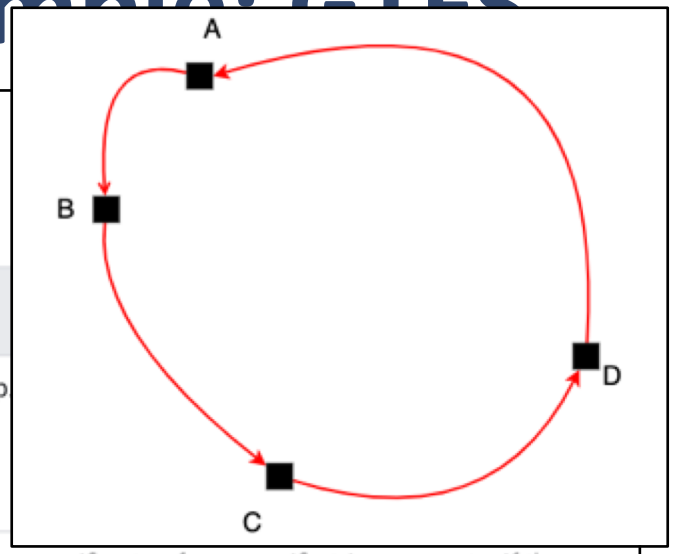
- Data is siloed: acquired and generated data is expensive, but often not reused
- Multitude of transportation planning tools are in use by researchers and cities
- No easy way to compare results as each has their own unique data models

A standard for this data is needed!

What about existing standards?

- Scope: existing standards overlap with, but **don't cover the domain of transportation planning**.
- Encoding limitations: traditional standards are **subject to ambiguity**, despite detailed definitions

Ambiguity Example: GTFS



stop_times.txt

File: **Required**

Field Name	Type	Required	Description
`trip_id`	ID referencing `trips.trip_id`	**Required**	Identifies a trip
`arrival`	Time	**Conditionally**	Arrival time at a specific stop for a specific trip on a route. If there

What about loop routes?

`stop_id`	ID referencing `stops.stop_id`	**Required**	Identifies the serviced stop. All stops serviced during a trip must have a record in stop_times.txt. Referenced locations must be stops, not stations or station entrances. A stop may be serviced multiple times in the same trip, and multiple trips and routes may service the same stop.
`stop_sequence`	Non-negative integer	**Required**	Order of stops for a particular trip. The values must increase along the trip but do not need to be consecutive. *Example: The first location on the trip could have a* `stop_sequence=1`*, the second location on the trip could have a* `stop_sequence=23`*, the third location could have a* `stop_sequence=40`*, and so on.*

- GTFS: General Transit Feed Specification[1]
- Common format for public transit data (schedules, locations,...)
- Highly successful, widely adopted

[1] https://developers.google.com/transit/gtfs

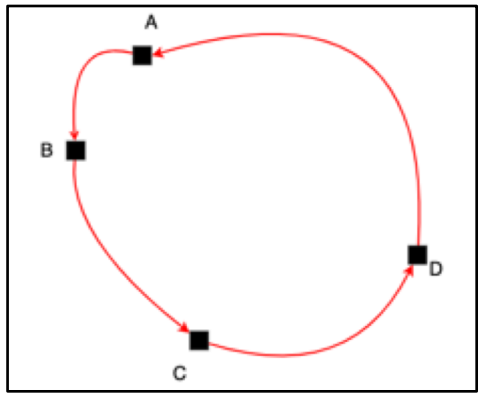
Example: How to define stop times for loop routes?

Option A:

Trip_id	arrival_time	departure_time	stop_id	stop_sequence	stop_headsign
trip_1	06:10:00	06:10:00	stop_A	1	"outbound"
trip_1	06:15:00	06:15:00	stop_B	2	"outbound"
trip_1	06:20:00	06:20:00	stop_C	3	”inbound"
trip_1	06:25:00	06:25:00	stop_D	4	"inbound"

Option B:

Trip_id	arrival_time	departure_time	stop_id	stop_sequence	stop_headsign
trip_1	06:10:00	06:10:00	stop_A	1	"outbound"
trip_1	06:15:00	06:15:00	stop_B	2	"outbound"
trip_1	06:20:00	06:20:00	stop_C	3	”inbound"
trip_1	06:25:00	06:25:00	stop_D	4	"inbound"
trip_1	06:30:00	06:30:00	stop_A	5	“”



- Not explicitly addressed in the reference
  - Open-ended!
- Identified as a special case in the GTFS best practices document
  - *Recommendation*: Option B

Ambiguity in Standards Specifications

Traditional approach

- Detailed documentation
- Modelling languages that focus on the data's structure as opposed to its semantics
- Meaning is grounded in natural language

LEADS TO

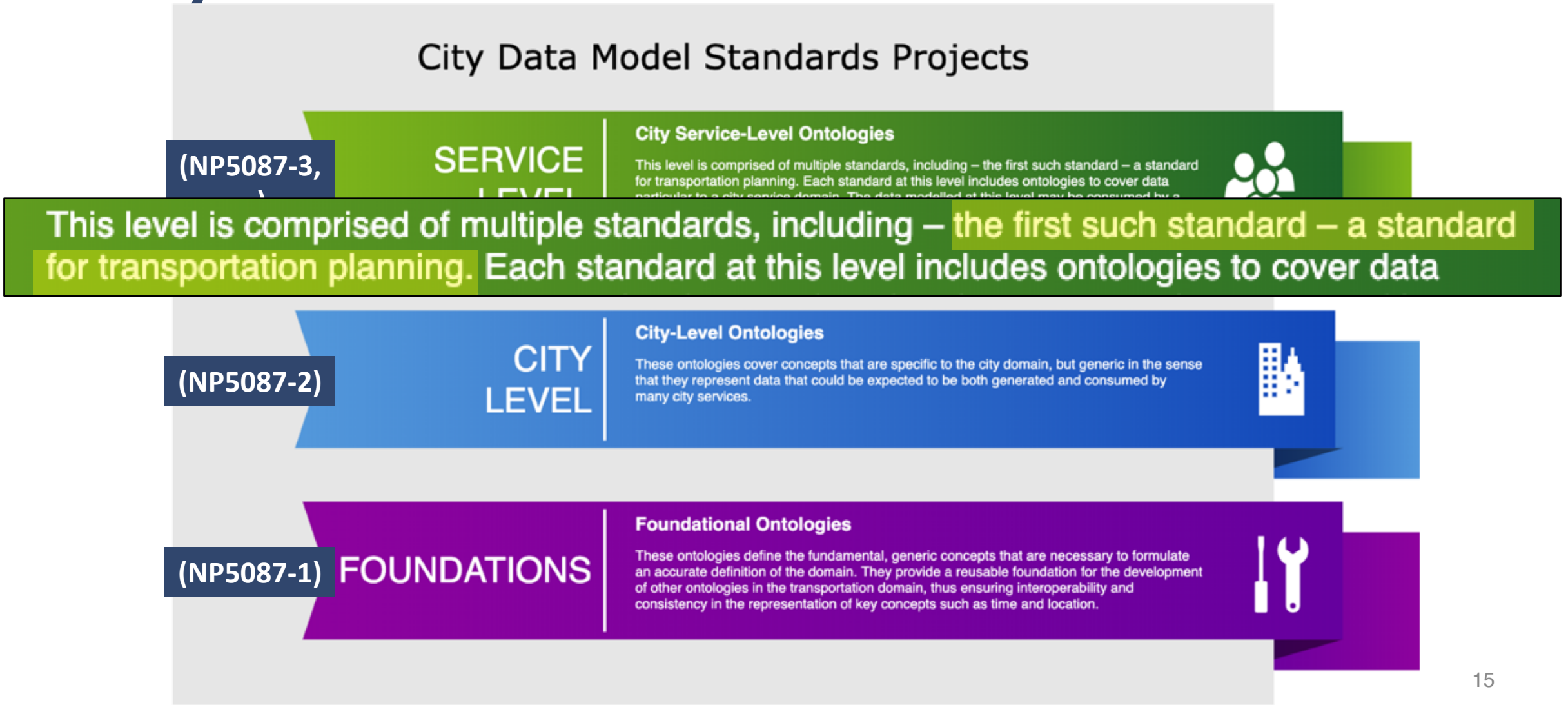
Challenges

- Natural language inherently ambiguous
- Need for supplementary material to resolve individual issues; clarify/recommend interpretations
  - Examples, best practices,…
  - But can't predict or detect them all
- Differences in interpretation lead to differences in adoption
  - Impacts the standard's effectiveness

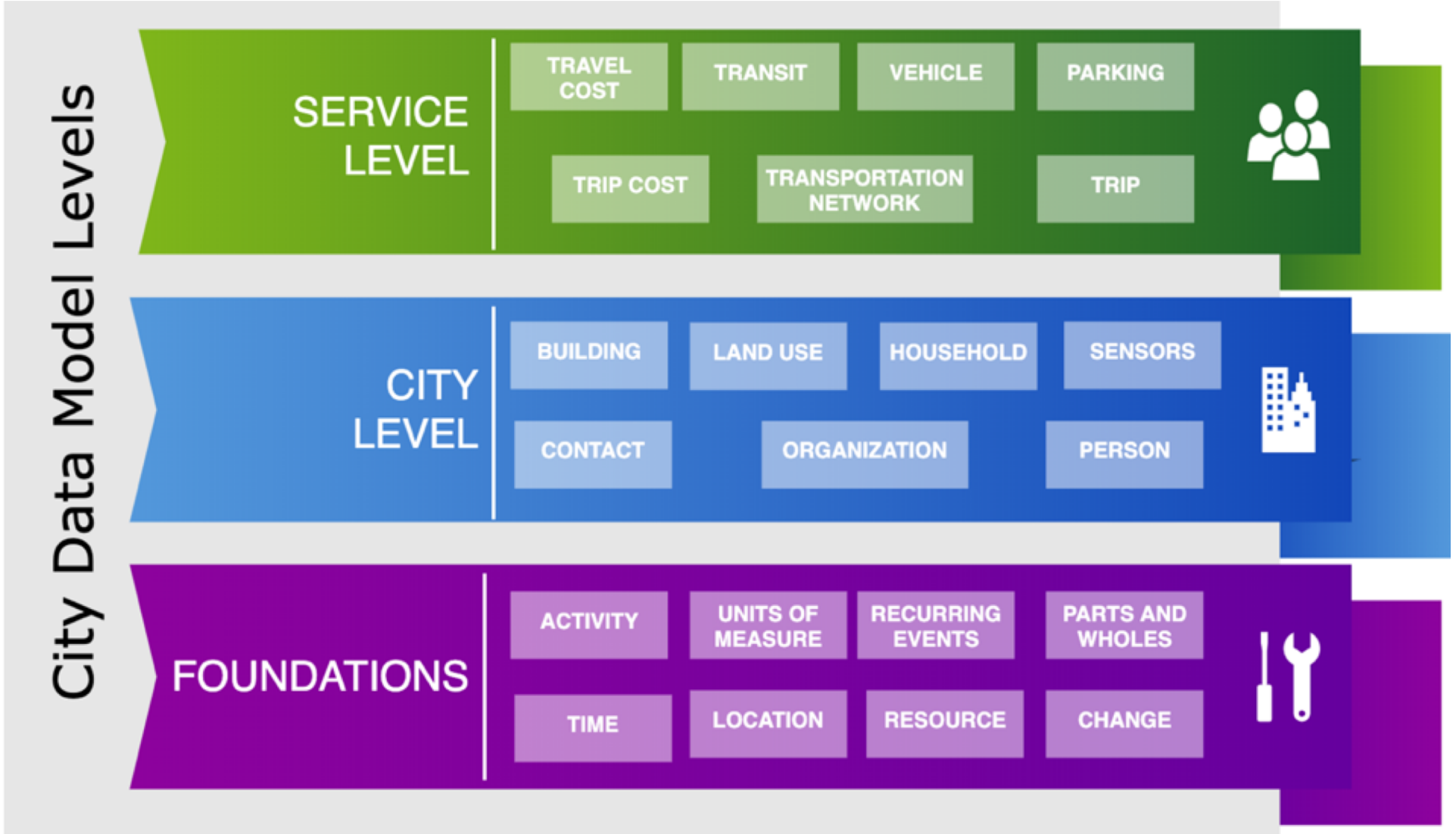
An ontology-based standard

- Has a **unique** interpretation:
  - Explicit, unambiguous encoding
  - Incorrect and correct interpretations may be automatically identified
- Added benefits:
  - Works with different tools and data formats
  - Supports a dynamic domain: core concepts are easily extended
  - May be implemented for other applications (e.g. reasoning)

ISO/IEC JTC1 WG11 Smart Cities
City Data Model NP5087



City Data Model: Transportation Planning (NP5087-3)



Standards Collaboration

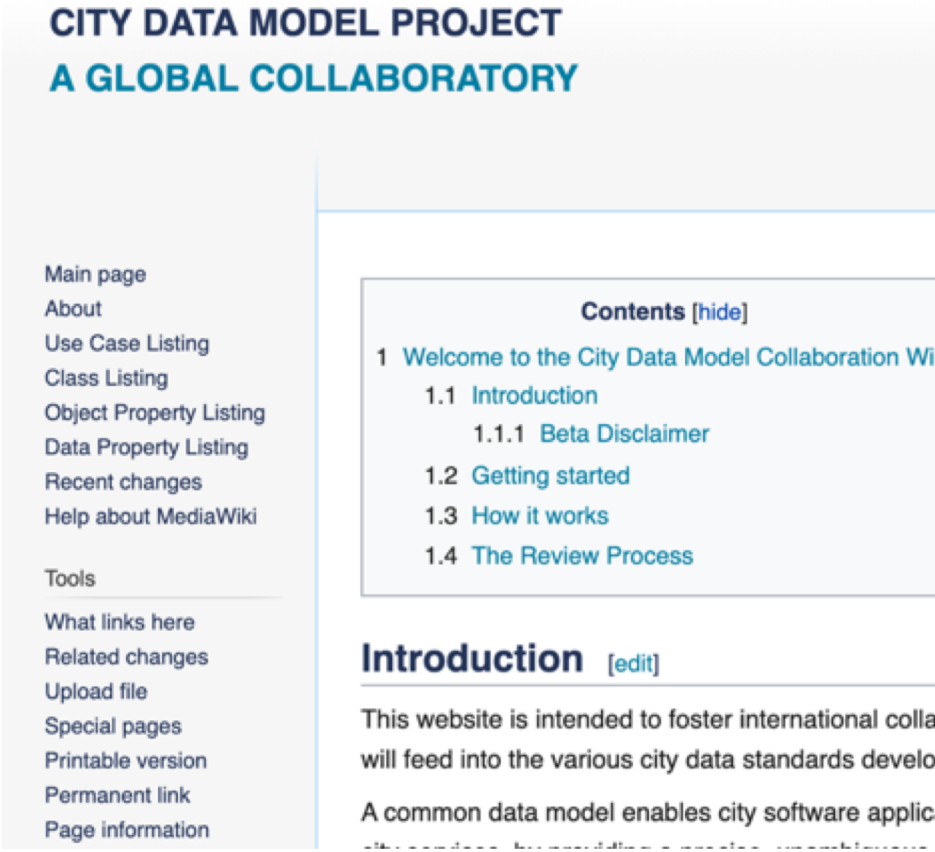
- Another standard, another silo?
  - How can we avoid this?
- There is a need to collaborate with other groups in order to understand how definitions of overlapping concepts are related
- We created a **Global Collaboratory** to support the alignment of the city data model standards with other standards efforts

The City Data Model Global Collaboratory

http://citydata.utoronto.ca [2]

- Develop a global consensus on the City Data Model
  - Identify concepts and definitions to be included
  - Align related concepts across standards

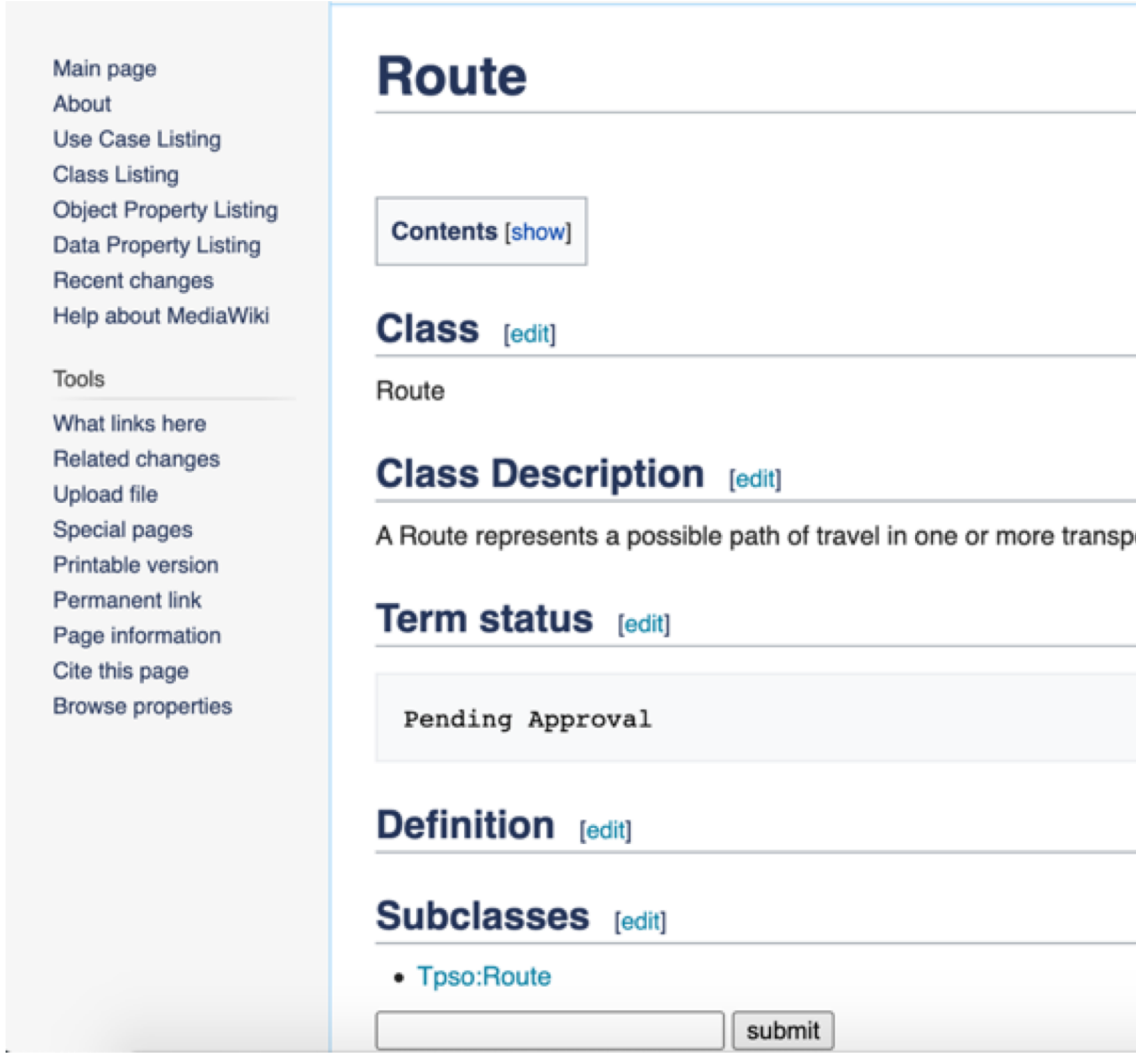
- Tasks supported:
  - Browse and review content
  - Comment on existing content and suggest changes or revisions
  - Propose terms and definitions
  - Submit use cases to explain/justify terms and definitions



[2] To be: citydatastandard.org

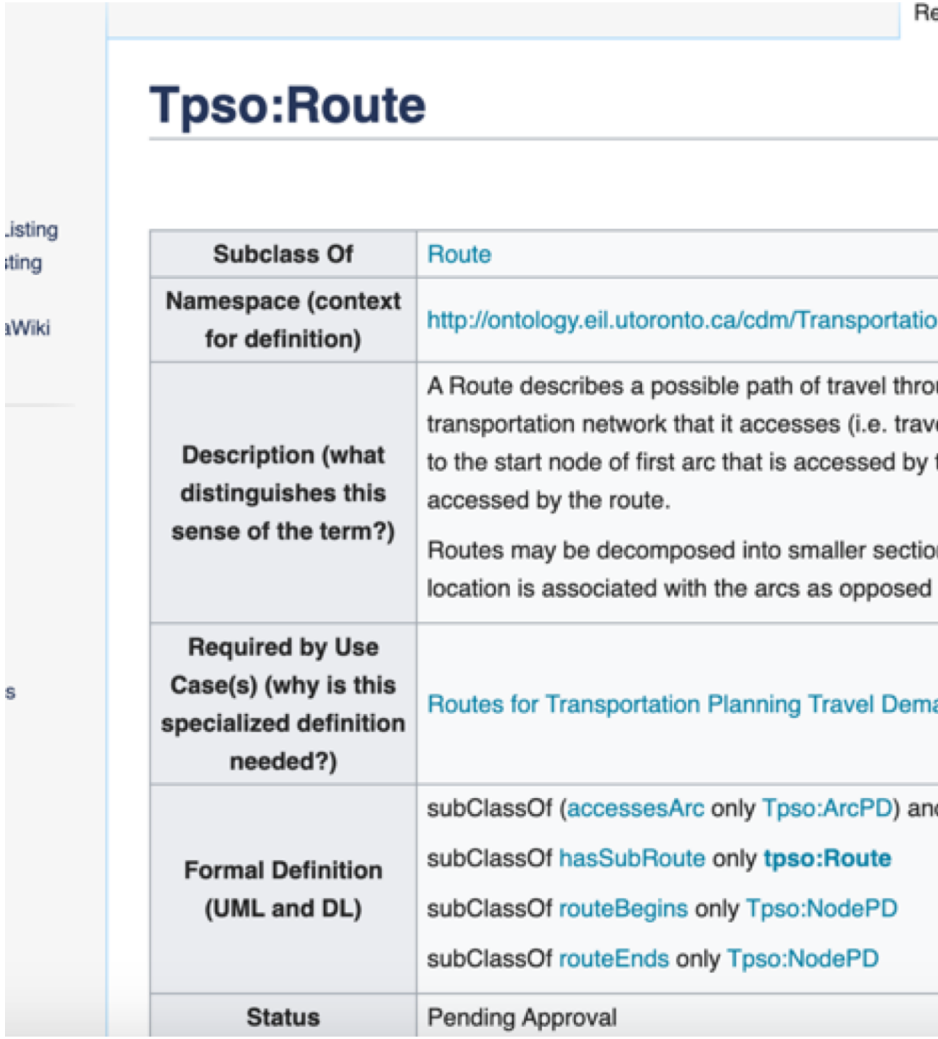
The Proposal Process: New terms

- Anyone can propose a new term.
  - Once a term has been proposed, it becomes open for the specification of definitions from the community.
- Terms may be independent, or proposed as specializations of existing terms (i.e. sub-classes or sub-properties)



The Proposal Process: New definitions

- Proposed definitions for a term must be specified in a *formal* language (Description Logic or UML)
- Proposed definitions must be accompanied by a use case
- Subject to community review
- Multiple definitions allowed and expected
  - Do not need to agree on one
  - Once discussion of definitions has settled, administrators will review all proposed definitions and identify the **minimum viable definition** for the term.

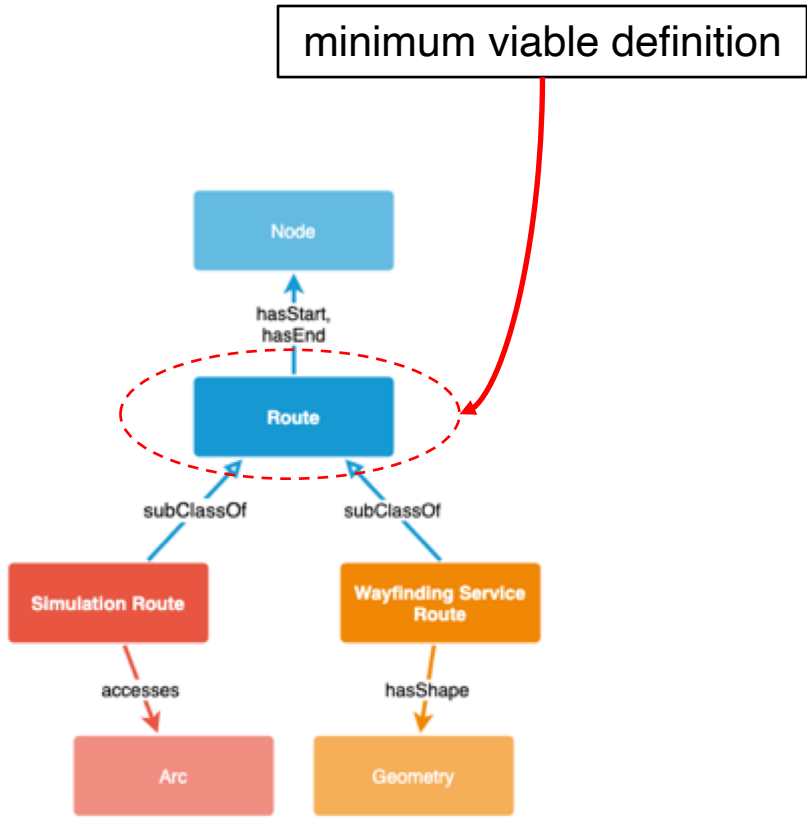


Tpso:Route

Subclass Of	Route
Namespace (context for definition)	http://ontology.eil.utoronto.ca/cdm/Transportatio
Description (what distinguishes this sense of the term?)	A Route describes a possible path of travel thro transportation network that it accesses (i.e. trav to the start node of first arc that is accessed by accessed by the route. Routes may be decomposed into smaller sectio location is associated with the arcs as opposed
Required by Use Case(s) (why is this specialized definition needed?)	Routes for Transportation Planning Travel Dema
Formal Definition (UML and DL)	subClassOf (accessesArc only Tpso:ArcPD) an subClassOf hasSubRoute only **tpso:Route** subClassOf routeBegins only Tpso:NodePD subClassOf routeEnds only Tpso:NodePD
Status	Pending Approval

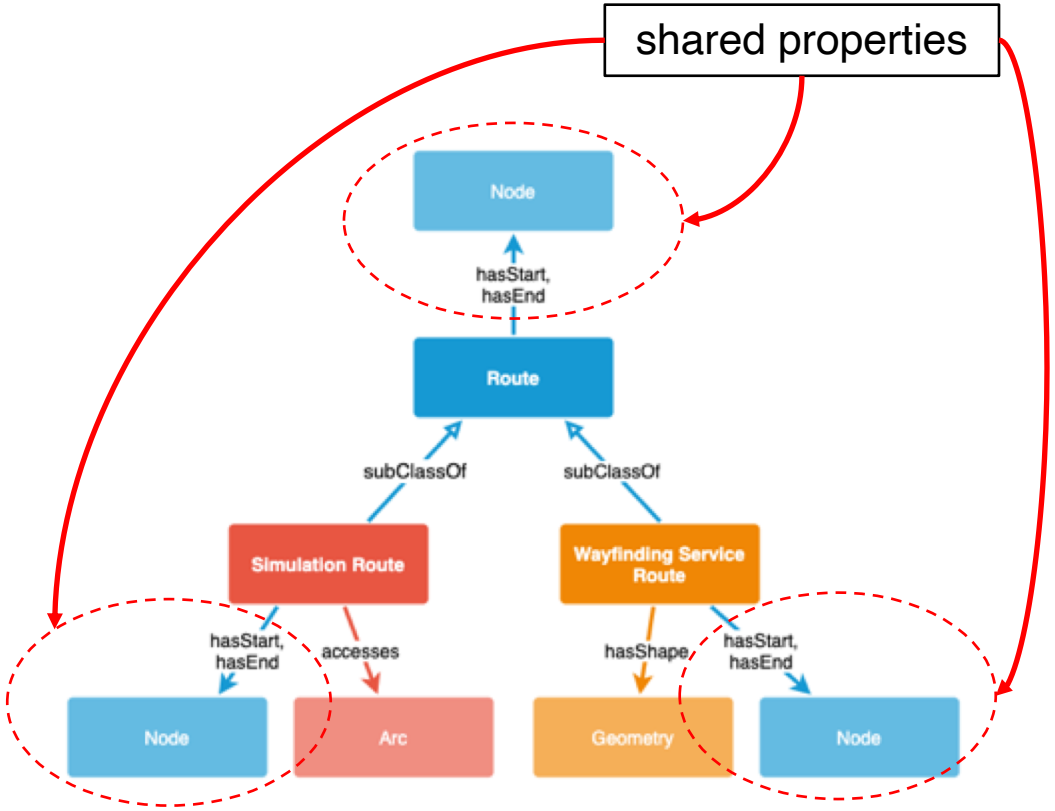
The Proposal Process: Minimum Viable Definitions

- Identified relative to a set of definitions for the same term
- The *minimum* semantics required for a term
  - Shown to be shared between each definition in the set



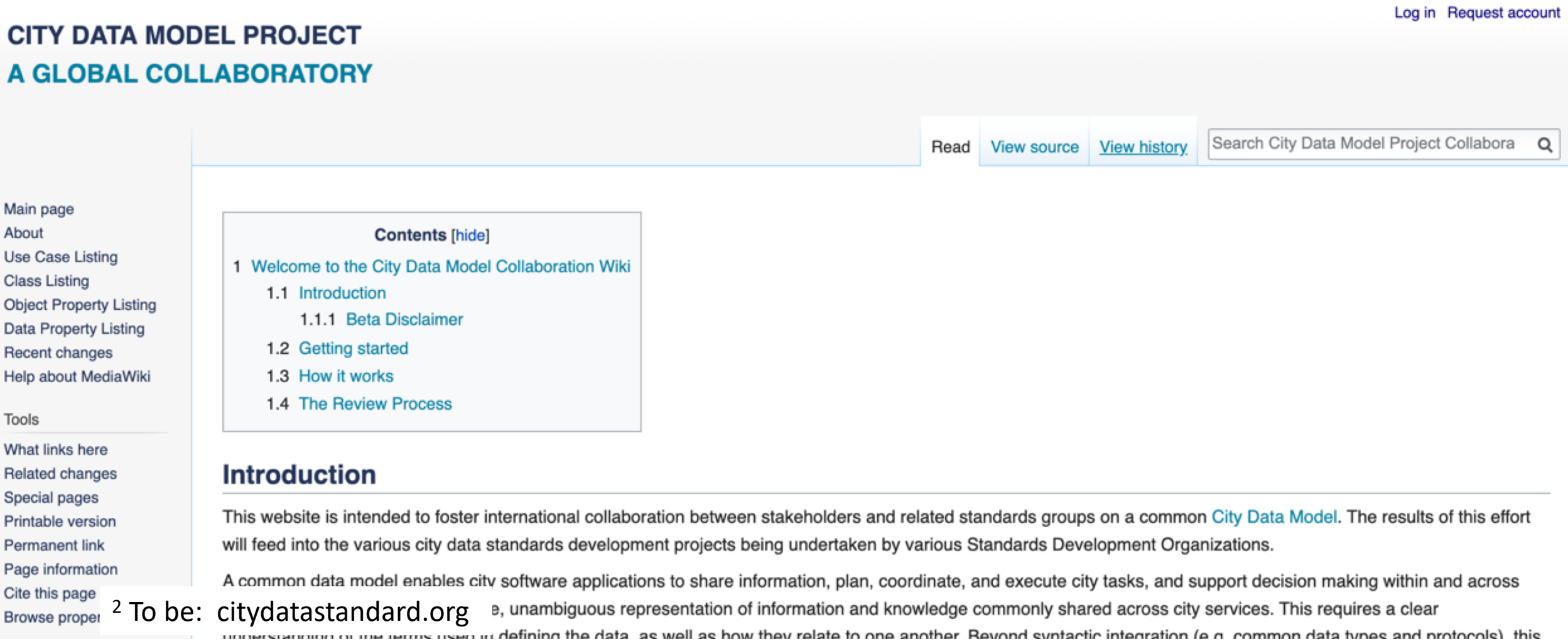
Minimum Viable Definitions

- Clearly identify shared meaning attributed to a particular term
  - Distinct definitions identified as subclasses of the minimum viable definition
  - Minimum viable definition identifies the set of properties shared amongst subclasses
    - Subset of shared terminology



Participants welcome!

http://citydata.utoronto.ca[2]



[2] To be: citydatastandard.org