An Linked Data Repository for Transportation Planning Data

Megan Katsumi, Mark Fox

Enterprise Integration Lab (eil.utoronto.ca)

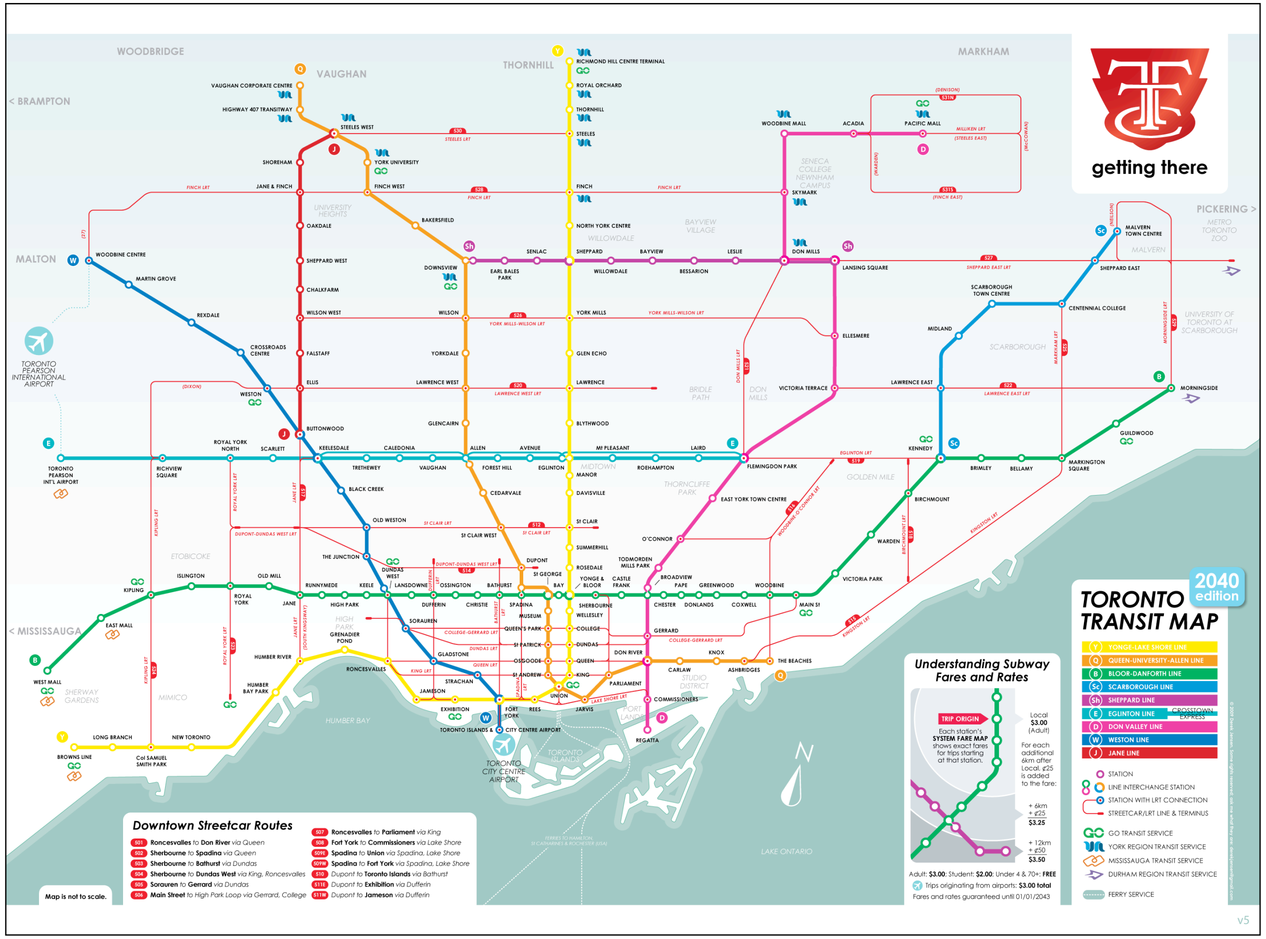
University of Toronto Transportation Research Institute (uttri.ca)

University of Toronto

Friday, May 31, 2019

Mechanical & Industrial Engineering
UNIVERSITY OF TORONTO

Transportation Planning

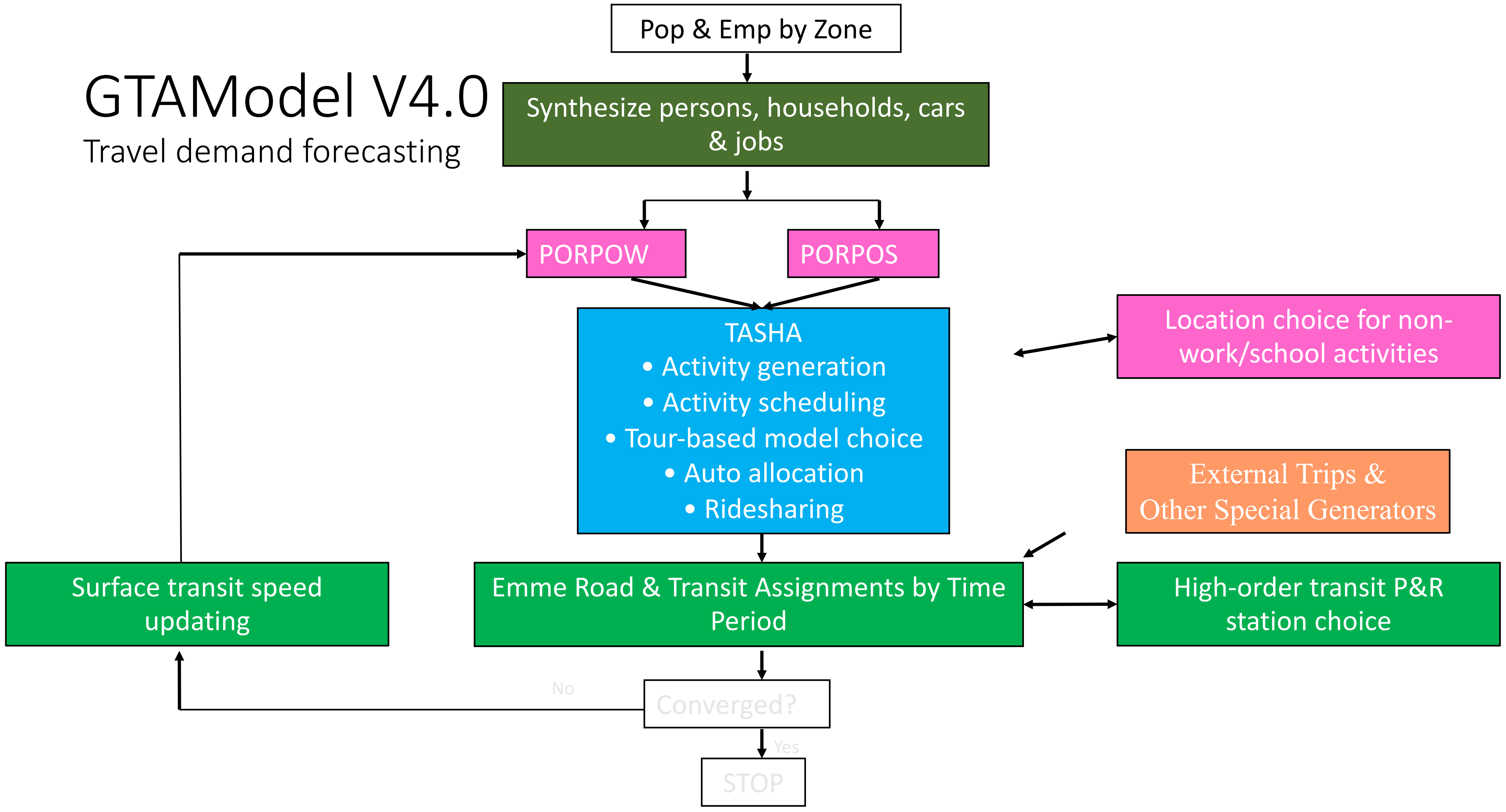


Problem: Planning transportation infrastructure over a 30 year horizon

- What will demand for public transportation and roads be over the next 30 years?
- How do changes in transportation infrastructure affect travelers?
- What are the environmental impacts of growth?

GTAModel V4.0

Travel demand forecasting



Challenges for transportation planning data

1. Data representation:
   - 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

   We need a standard for transportation planning data!

2. Data storage:
   - Heterogeneity of data: Data required to support transportation planning is available in different formats with different representations. Consequently, related data are often stored in isolation.
   - Wasted effort: Data that is cleaned and integrated for one task may not be reused for others

   We need a more effective way of storing transportation planning data!

Challenge #1: Data representation

- Problem:
  - Multitude of transportation planning tools are in use by researchers and cities
  - No easy way to compare or reuse results as each tool has its own unique data models

Solution: a standard for transportation planning data

- The problem may be addressed with a standard that:
  - Facilitates interoperability between heterogeneous data
    - Works with different tools, data formats
  - Is easily extensible: tools and approaches are always changing
  - Has a **unique** interpretation; incorrect and correct interpretations should be clearly identifiable
- **Claim:** an ontology can be used to specify a standard that will satisfy these requirements

Transportation Standards for Interoperability

- Data Standards
  - Transmodel
  - gtfs
  - ISO 19107 Road network model
  - ISO 14825 Geographic Data Files (GDF) standard
  - Geographic Markup Language (GML)
  - DATEX II
  - W3C Vehicle Information Service Specification
  - SENSORIS
  - ADASIS
  - …

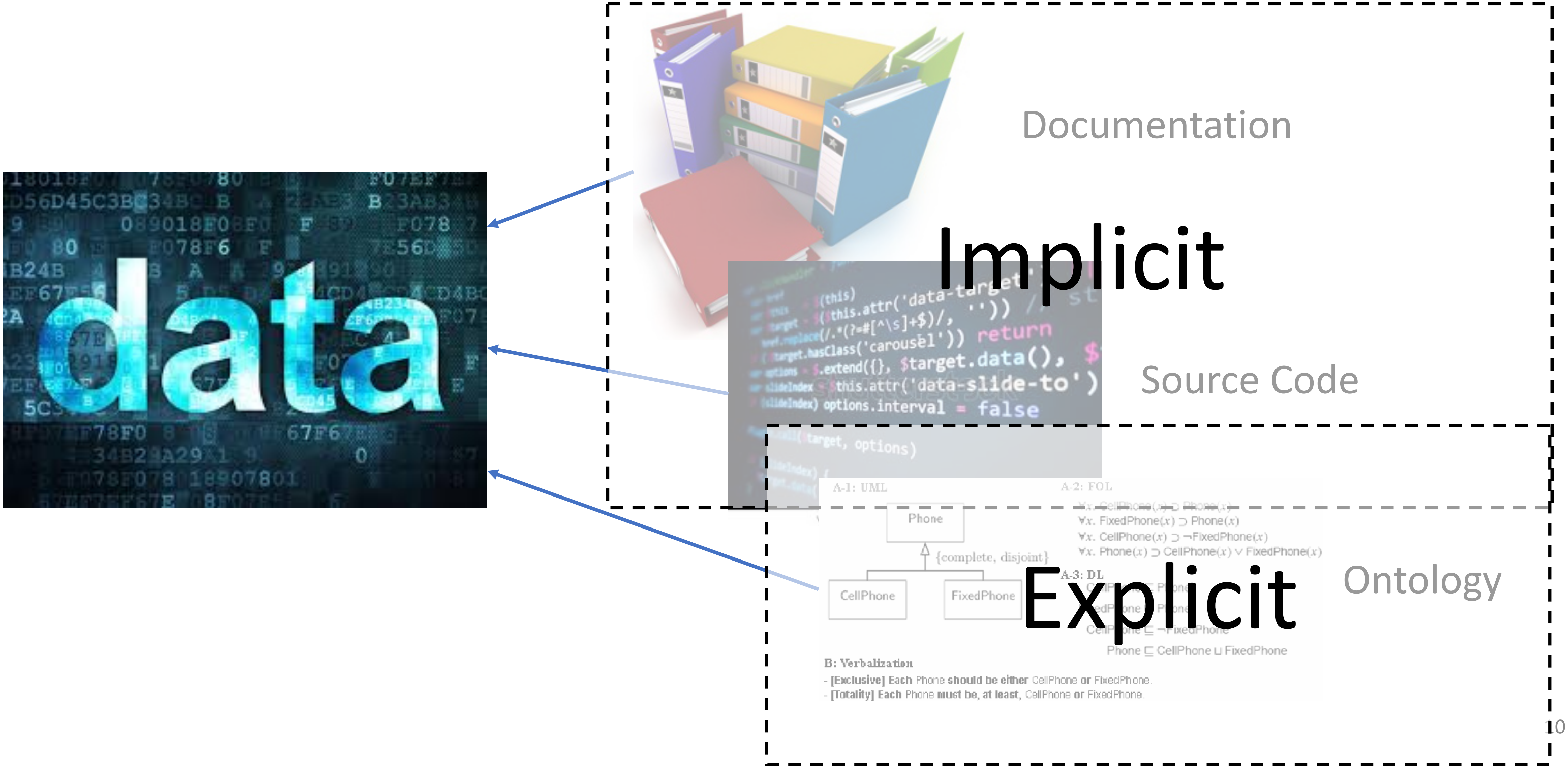
Existing data standards

- Don’t focus on transportation planning
- Subject to ambiguity
  - Potential for misinterpretation
  - Correspondences with other standards unclear
- May support syntactic interoperability, but cannot support *semantic* interoperability

What is Semantic Interoperability?

- The ability of computer systems to exchange data with unambiguous, shared meaning.
  - A requirement to enable machine computable logic, inferencing, knowledge discovery, and data federation between information systems.
- Is concerned not just with the packaging of data (**syntax**), but the simultaneous transmission of the meaning with the data (**semantics**)

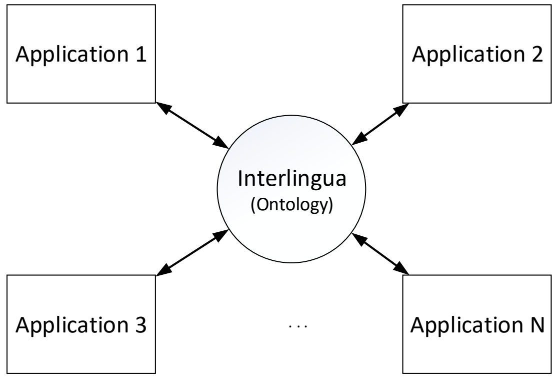
Sources of Data Semantics



What is an Ontology?

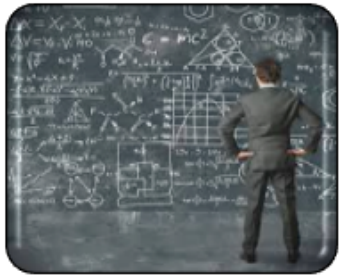
- (More than) a reference model for the domain.
- Answers the questions:
  - What are the core concepts and properties that span the city's data?
    - To what extent can we generalize them in a useful way?
  - What are the key distinctions?
    - Can we formally define necessary and/or sufficient conditions (using properties) for something to be an example (member) of a concept?
- A precise, formal (logical language) representation that supports:
  - Reuse
  - Integration
  - Automated deduction

How are Ontologies Used?



- Data Integration:
  - Ontology to serve as an *interlingua*
  - Data and systems may be mapped into the ontology to support exchange of information
- Automated Deduction
  - New information may be inferred based on the data and knowledge of the domain formalized with the ontology.
- Model Checking:
  - Data may be automatically validated against the ontology to check whether it conforms to the definitions.

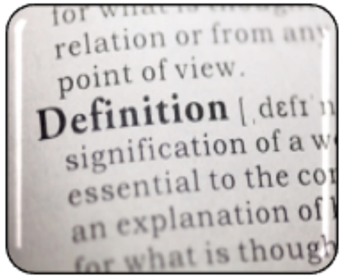
Ontology Components



Micro-Theory

- Axioms/Rules
- Deduction – answering questions

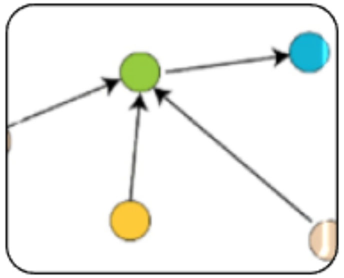
- For each year above the age of 14, a member of a household will leave with a probability p(Age)



Definitions and Constraints

- Class Definitions (in Logic)
- Automated classification

- Household is composed of at least one person who resides at the same address



Knowledge Graph

- Classes and Properties
- Taxonomy and Inheritance

- Households
- Transportation Network
- Vehicles

Example: City Resident

- **Toronto**: “you are identified as a resident if you reside in, own property, or own or operate a business in Toronto” (311 Toronto).
- **Beijing**: “all individuals holding the nationality of the People’s Republic of China who [have] a domicile in Beijing and nowhere else. If the individual maintains a regular dwelling somewhere else, the more regular dwelling is considered their place of residence” (Li, 1991).
- **New York**: “the place which an individual intends to be his permanent home – the place to which he intends to return. It is the home with range of sentiment, feeling and permanent association. One must be domiciled in New York and maintain a home in New York, the time spent in the State is irrelevant” (McGladrey, 2009).
- **Germany**: “a resident of Germany generally refers to an individual who has a domicile in Germany or spends more than six consecutive months in Germany (habitual place of abode)” (Seidel, 2011).

How to Express the Semantics of Resident?

“you are identified as a resident if you reside in, own property, or own or operate a business in Toronto”

TorontoResident is-a Resident

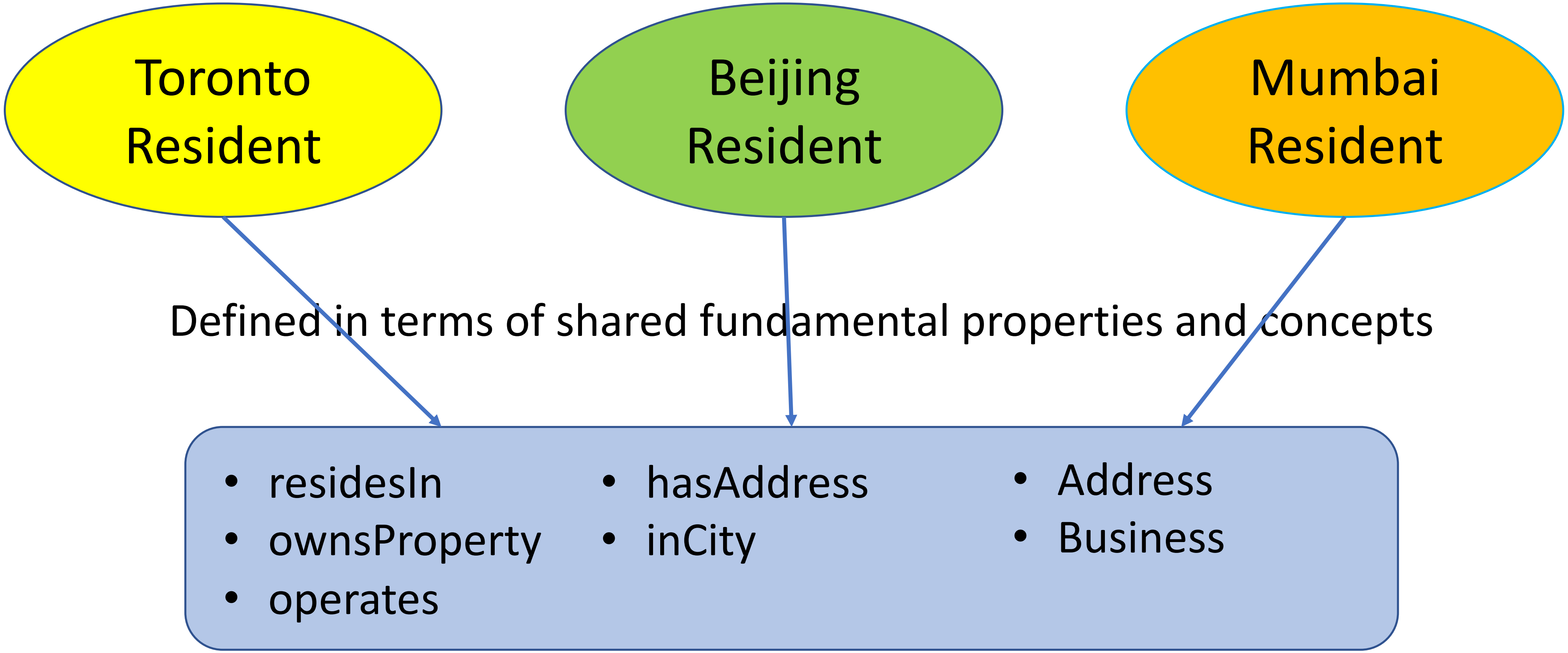
and (residesIn.Toronto

or ownsPropertyIn.Toronto

or operates.(Business

and hasAddress.(Address and inCity.Toronto)))

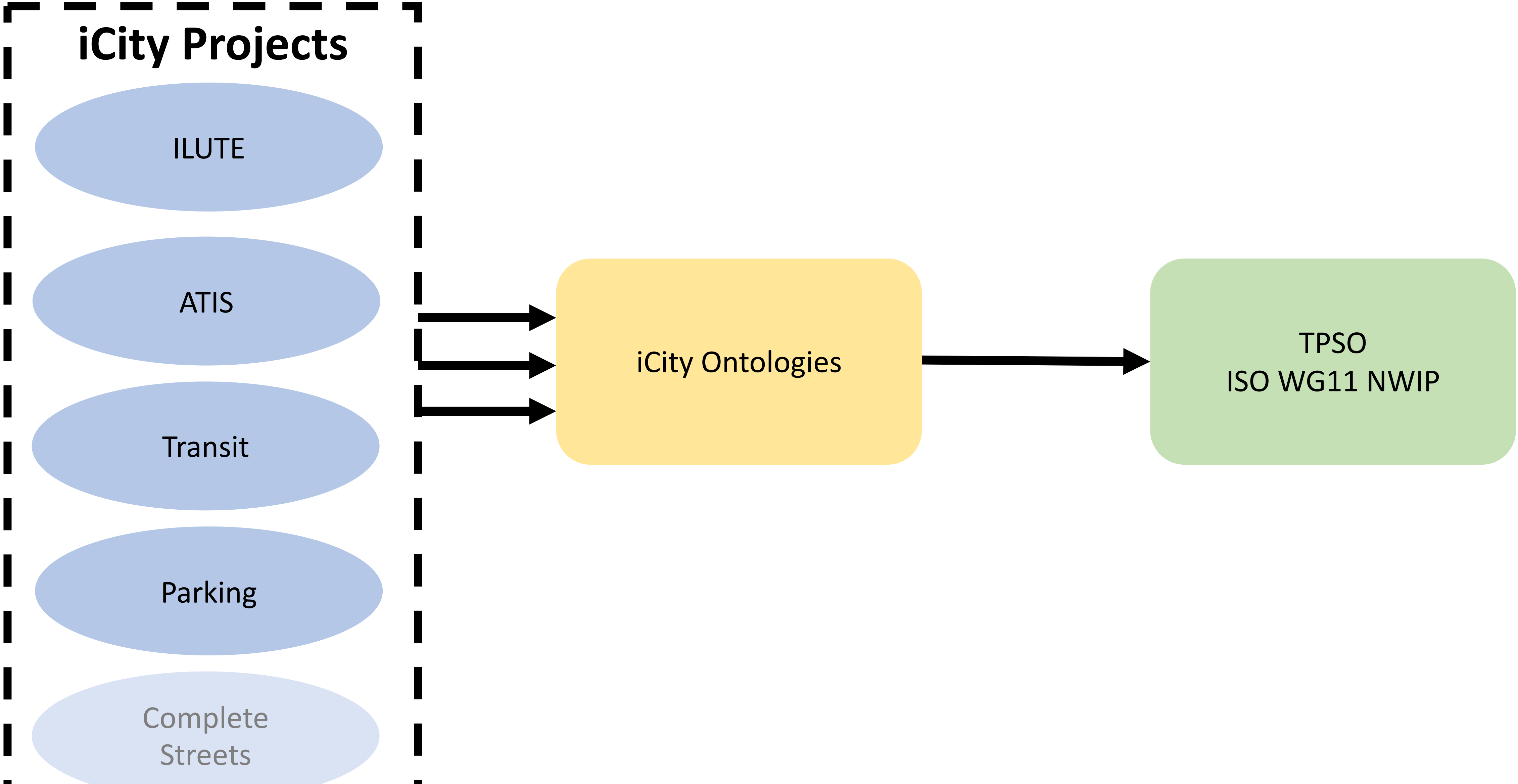
Different Views of a Concept



An ontology-based standard for transportation planning data

- Requirements
  - ✓Facilitate interoperability between heterogeneous data
  - ✓Must work with different tools, data formats
  - ✓Must have a **unique** interpretation; incorrect and correct interpretations should be clearly identifiable
  - ✓Must be easily extensible: tools and approaches are always changing

Ontologies for Transportation Data



TPSO: Components

Transportation Network
Transit
Vehicle
Travel Cost
Trip Cost

Household
Person
Organization
Trip
Contact

Building
Land Use
Parking

Location
Time
Units of Measure
Change
Activity
Sensors

Solution for Challenge #1

- Use **ontologies** to specify a standard for transportation planning data
  - Project 1.1 has been developing an ontology for the iCity project
  - This ontology is being applied to serve as the basis for a standard for transportation planning (the TPSO) with ISO WG11

Challenge #2: Transportation Data Storage

- Problem:
  - Heterogeneity of data: Data required to support transportation planning is available in different formats with different representations. Consequently, related data are often stored in isolation.
  - Wasted effort: Data that is cleaned and integrated for one task may not be reusable for others

Solution: a more effective data repository

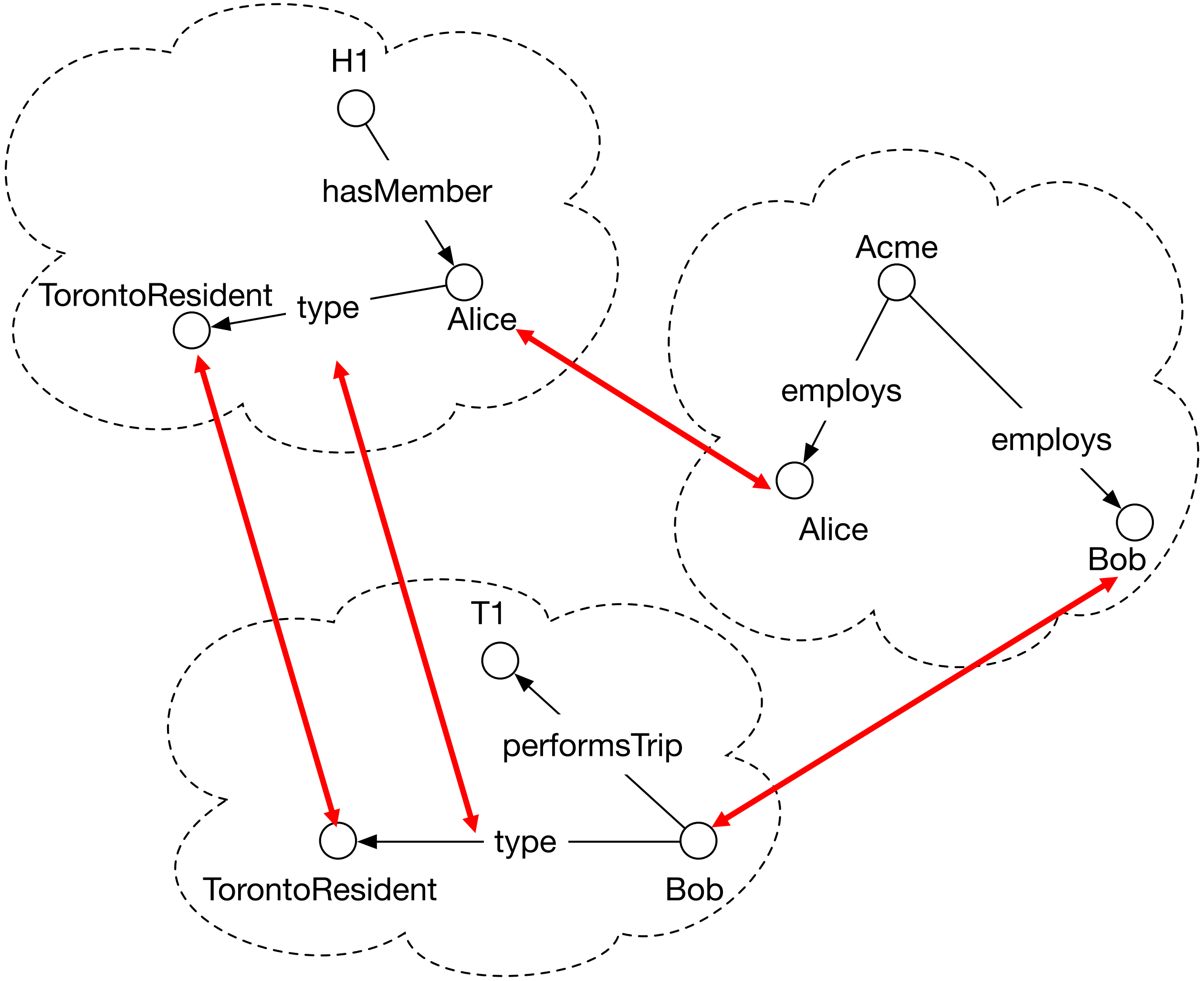
- The challenges for transportation data storage may be addressed with a repository that:
  - Interprets data semantics in a clearly defined, commonly agreed upon, and unambiguous way
  - Captures and leverages the relationships between data
  - Tracks data provenance
- **Claim:** A linked data repository can satisfy these requirements

Linked Data Repositories

Linked data repositories store datasets using semantic web technologies

- Data is captured using a standard language (RDF)
- A shared vocabulary, defined with an ontology, may be used to provide semantics for the data
- Things (vocabulary terms and instances) are identified uniquely, using Uniform Resource Identifiers (URIs)
  - This creates the ability to link within and between datasets
  - Linked data can be queried using the terms defined in the ontology

Linked Data Repositories



Traditional vs Linked Data Repositories

Feature	Traditional Repository	Linked Data Repository
Support different formats and schema	With preprocessing	With preprocessing; some direct translation possible
Clear interpretation of data semantics	✗	✓
Easily extensible	✗	✓
Capture the relationships between data	Sometimes*	✓
Track provenance	Sometimes	✓
Reason about provenance	✗	✓
Easy access to data	Sometimes*	✓
Data validation (QA)	Sometimes*	✓
Perform reasoning with the data	✗	✓

*If a central DB with a well-defined schema is used to store the data

Solution for Challenge #2

- Use ontologies to represent the semantics of the data
  - The proposed TPSO standard may be leveraged to facilitate correct integration and reuse of data.
- Use a linked data repository and the ontologies to store transportation planning data
  - Ongoing work (Year 5)
  - GUDR: the Global Urban Data Repository is designed to serve as a linked data repository for all urban data

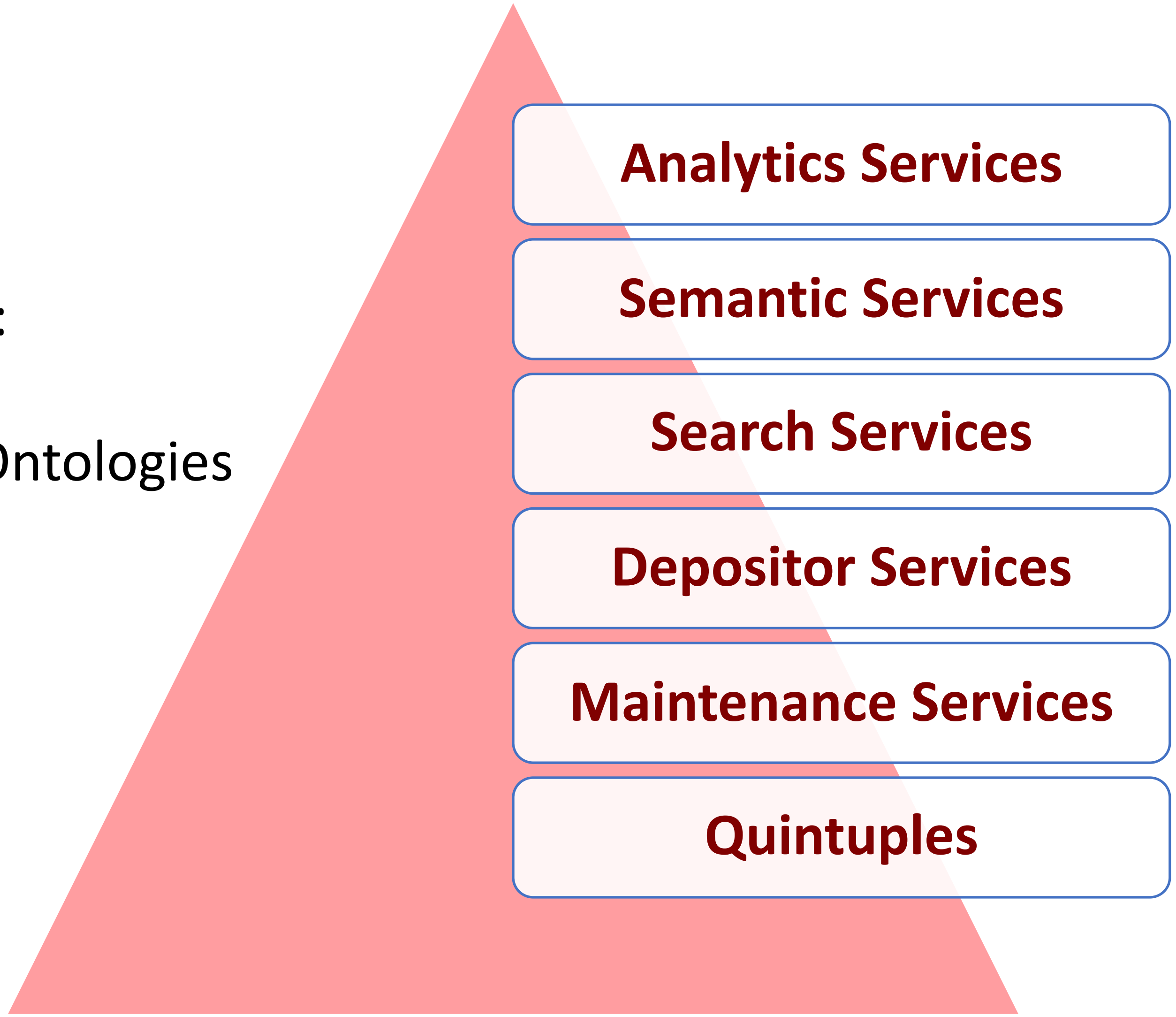
GUDR: Global Urban Data Repository

“A linked data repository in which all urban data can be deposited, searched and retrieved.”

GUDR Stack

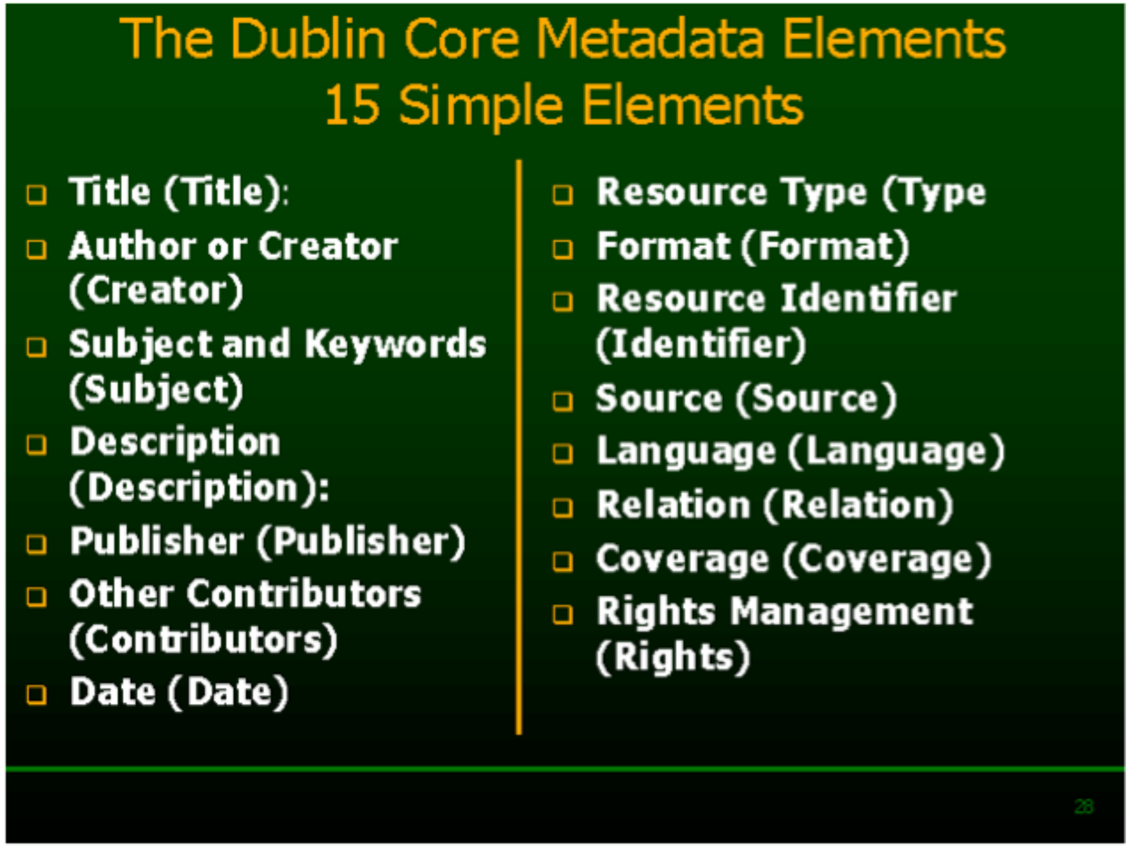
Key Characteristics:

1. Meta-Data
2. Vocabularies/Ontologies



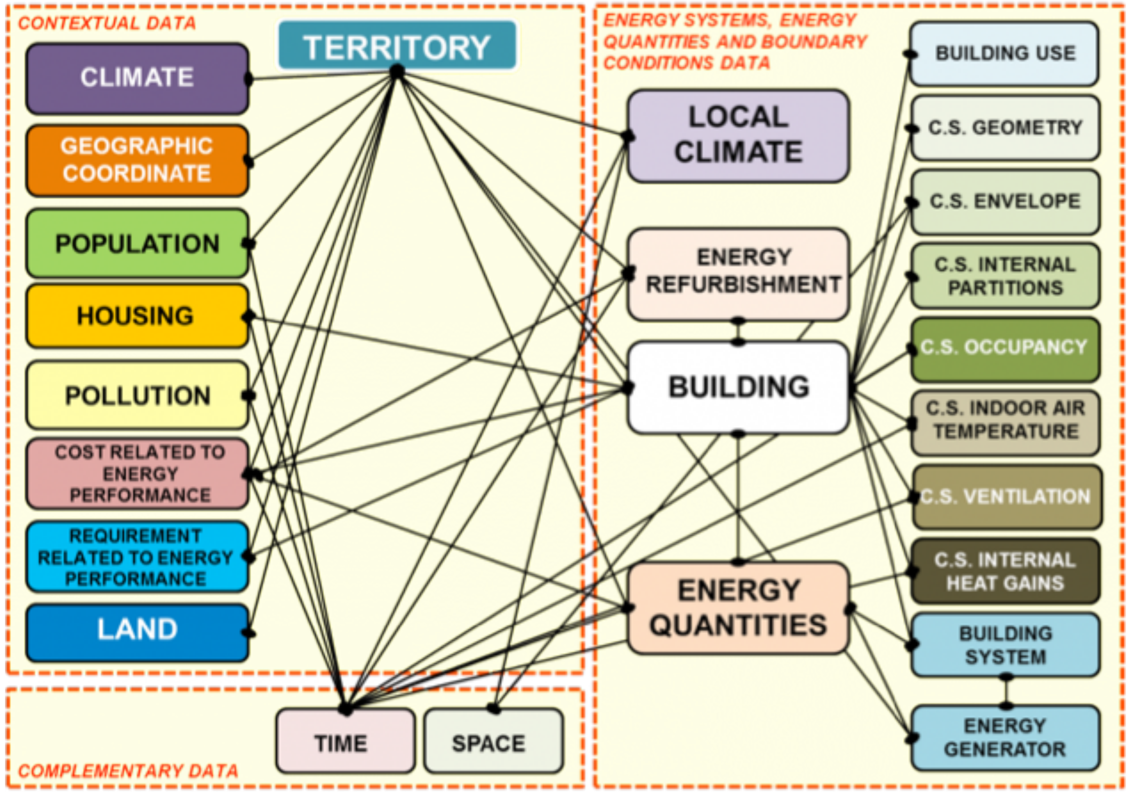
Meta-Data

- Each RDF triples has meta-data attached to it.
- The process of depositing automatically attaches meta-data.



Vocabularies/Ontologies

- The repository includes vocabularies/ontologies.
- Depositors choose which to use.



Next Steps

- Standardization of TPSO with ISO/JTC 1 WG 11 Smart Cities an ongoing effort
- Continue development and implementation of GUDR
  - Ongoing development and research, e.g. data validity
- Leverage the TPSO and GUDR to provide a linked data repository for transportation planning data

Questions?