

Reliable Foundations For Asset Management

Data Strategy and Excellence

Jasmine Joy



- Data Engineer, BGIS Canada
- BSc. Physics, Computer Science
- Certificates: GIS, ML
- GIS, Data, Business Analyst
- Strengths: pipelines, governance, management

Agenda



Asset Hierarchy



**Asset Data
Model**



Data Dictionary



**Data Governance
Frameworks**



Questions



Asset Hierarchy



**Asset Data
Model**



Data Dictionary



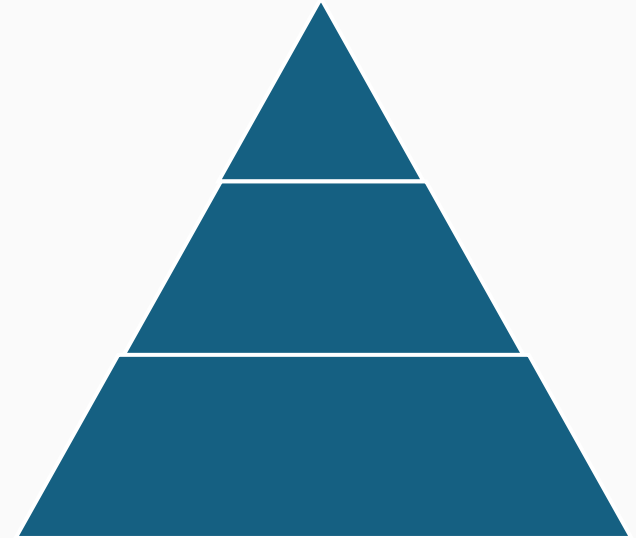
**Data Governance
Frameworks**



Questions

Asset Hierarchy

- Efficient, organized list of equipment.
- Roadmap to capture data elements.
- Workshop with stakeholders.
- Understand existing processes.
Standardize.
- Define clear ownership categories,
reporting structure, maintenance
schedules.



Asset Hierarchy

1	Assets
2	— Linear (lines)
3	— Sidewalks
4	— Trails
5	— IrrigationPiping
6	— Electrical
7	— WaterLines
8	— Roads
9	— Fencing
10	— Green (polygons)
11	— Parks
12	— ForestAndLandCover
13	— Wetlands
14	— GreenSpaces
15	— FixedStructures (points)
16	— IrrigationSystems
17	— ParkingLots
18	— Buildings
19	— CommunityCenters
20	— TennisCourts
21	— FerryTicketCounters
22	— IceRinks
23	— Mobile (movables)
24	— Vehicles
25	— Lawnmowers
26	— Snowplow
27	— MaintenanceTrucks
28	— Amenity (points, not buildings)
29	— Benches
30	— TrashBins
31	— LightingPoles
32	— Trees
33	— PlaygroundEquipment
34	— WaterFountains
35	— Digital (points, not buildings)
36	— SurveillanceCameras
37	— LightSensors

- City of Toronto, Open Data Portal
- Consolidated “Parks, Forestry & Recreation” datasets.
- Used existing asset data, expanded list.
- *Capturing equipment monitoring data (time series).*
- Tools: Notepad, ETL/ELT, Python.
- Key components: **scalable & modular.**



Asset Hierarchy



**Asset Data
Model**



Data Dictionary



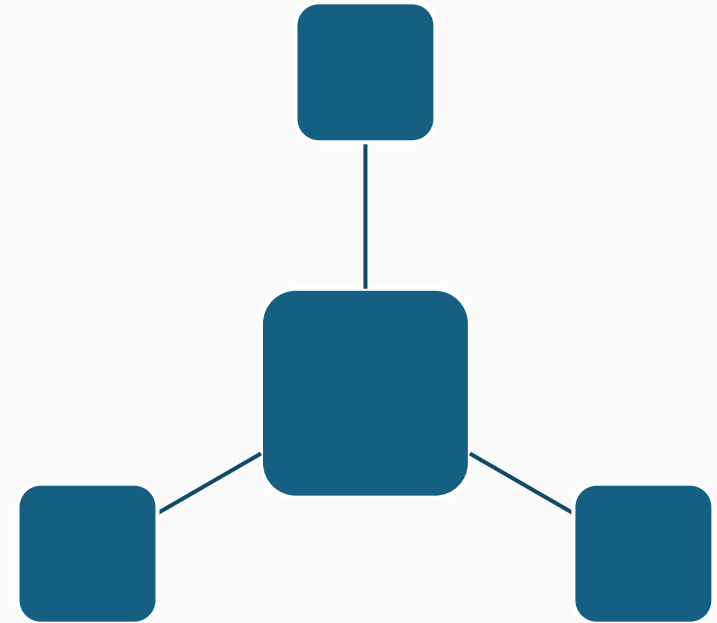
**Data Governance
Frameworks**



Questions

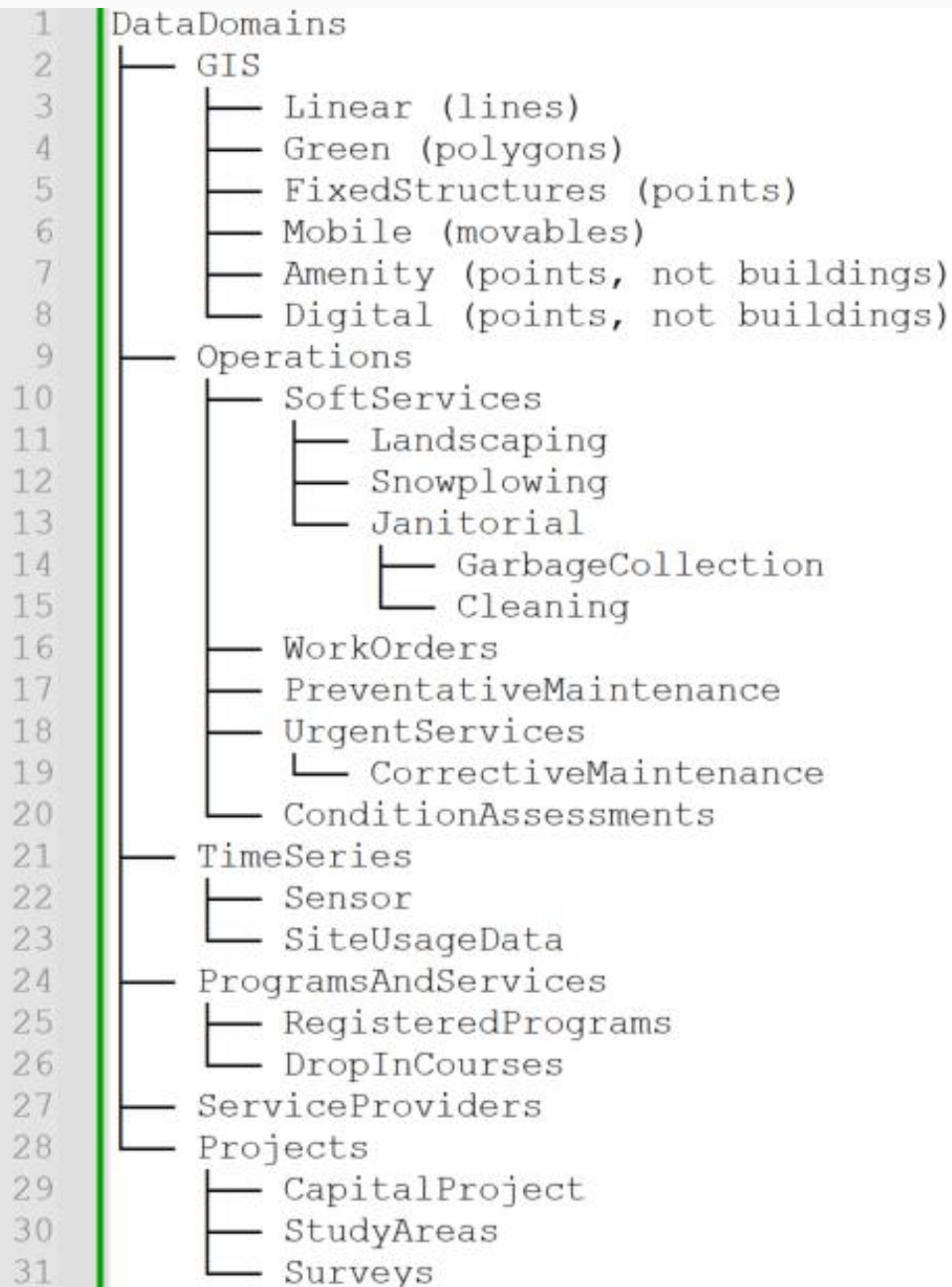
Asset Data Model

- Group by similar data components.
- Capture all necessary attributes (critical data elements).
- Dimensional modeling – STAR schema.
- Design with future use in mind.

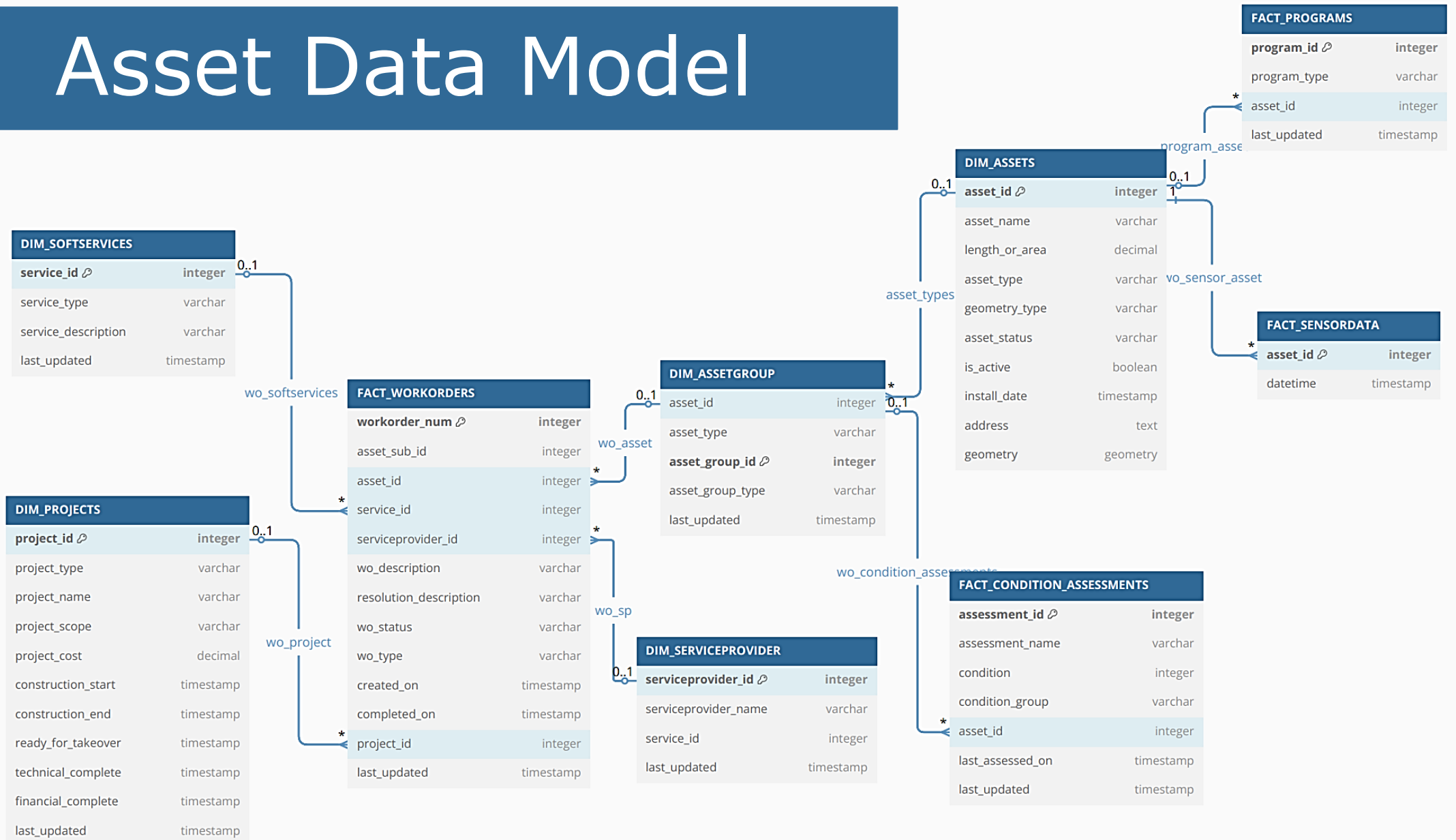


Asset Data Model

- *Vendor management data, API,*
 - *normalized,*
 - *then denormalized to avoid rework.*
- Tools: ETL/ELT, Python, dbdiagram.io
- Key components: **flexibility, maintainability & business relevance.**



Asset Data Model



Asset Data Model

1	DataDomains
2	GIS
3	Linear (lines)
4	Green (polygons)
5	FixedStructures (points)
6	Mobile (movables)
7	Amenity (points, not buildings)
8	Digital (points, not buildings)

DIM_ASSETGROUP	
asset_id	integer
asset_type	varchar
asset_group_id	integer
asset_group_type	varchar
last_updated	timestamp

DIM_ASSETS	
asset_id	integer
asset_name	varchar
length_or_area	decimal
asset_type	varchar
geometry_type	varchar
asset_status	varchar
is_active	boolean
install_date	timestamp
address	text
geometry	geometry

asset_types



Asset Hierarchy



**Asset Data
Model**



Data Dictionary



**Data Governance
Frameworks**



Questions

Data Dictionary

- Cornerstone for transparency and onboarding.
- Contain all fields, definitions.
- Alongside data model development.
- *Vendor management & Equipment monitoring.*
- Tools: Excel (for initial builds), Power BI (interactive), Git (version control).
- Key components: **quality, accessibility, maintainability.**

Database	Schema	Table	Field	Type	Description	Example
DB	DIM	ASSETS	asset_id	integer	Unique identifier for each asset.	SIDEWALK01
DB	DIM	ASSETS	asset_name	varchar	Commonly known name of the asset.	Sidewalk1
DB	DIM	ASSETS	asset_desc	varchar	Description to identify the asset.	Sidewalk on north side of Park
DB	DIM	ASSETS	length_or_area	decimal	Not relevant for points. Length (for linear assets (km)) or area (for polygons (sq	3
DB	DIM	ASSETS	asset_type	varchar	Specific classification, e.g., Building > Community Centers.	Sidewalk
DB	DIM	ASSETS	geometry_type	varchar	Type of geometry: Point, Line, or Polygon.	line
DB	DIM	ASSETS	asset_status	varchar	Current status of the asset (e.g., Operational, Decommissioned).	operational
DB	DIM	ASSETS	is_active	boolean	Indicates if the asset is active.	TRUE
DB	DIM	ASSETS	install_date	timestamp	Date the asset was installed, created, or acquired.	01-01-1990



Asset Hierarchy



**Asset Data
Model**



Data Dictionary



**Data Governance
Frameworks**



Questions

Data Governance Frameworks



Data Migration
Challenges



Data Quality



Testers



Documentation



Maintainability



Data Migration Challenges

- Text cleaning..... French/English, Montreal/Montréal
- Formatting issues in Excel..... Montréal , 01/12/2016 → 42705
- Data types & missing values..... '97', 97, 97.002
- Storage capacity..... Taking a reading every millisecond
- Incorrect Lat-Long.....

✓ *Bounding Box check with Geocoding*





Data Quality

- Data Quality metrics
- Validity best practice: only define failures.
 - ***"This field fails when..."***
- Score metrics.

Overall score

87%

Grade

Gold

Asset	Park Name	Parking Spots	Handicap Parking Spots	GIS Coordinate	Access
49503	BLUEHAVEN PARK			(-79.5563455530346,43.7449479239744)	Public
44683	BOOTH YARD			(-79.3431626237552,43.6557206841609)	Staff
44682	BOOTH YARD			(-79.3443166919273,43.6563295308317)	Staff
32230	BUDAPEST PARK			(-79.4468165517767,43.6366925316587)	Private
45163	CORONATION PARK - ETOBICOKE			(-79.4994743330933,43.6160145907673)	
32294	EGLINTON PARK		1	(-79.4039084226439,43.7059649849856)	Public
49520	FINCH - HUMBERLINE SPORTSFIELDS			(-79.6177210175333,43.7330823637425)	Private
44889	GARRISON COMMON			(-79.4053142139057,43.6382001513151)	Staff
32326	GUILD PARK AND GARDENS			(-79.1898013279774,43.748992471862)	Public
49517	HARRISON PROPERTIES				
32339	HIGH PARK				
49519	HIGHVIEW PARK		3		
45142	JEFF HEALEY PARK			(-79.4955596986474,43.6310386332859)	

City of Toronto: Parking Lot Facilities

Overall score

30%

Grade

Bronze

- ✓ Access: null -> 'Unknown'
- ✓ 'Parking Spots' > 'Handicap Parking Spots'
- ✓ GIS: Points within Toronto polygon
- ✓ Readme.txt contains recommended projection.



Testers

- Simplest test – is the data in today?
- Data type checks
- Data Migration:
 - Null value check
 - Random sampling
 - Row-count comparison
 - **Referential integrity test**
 - Hash comparison, table difference
 - Hash test, data drifts in time series
- *Developed data deficiency recognition system using ETL tools.*
- Tools: Unit tests in SQL(Stored Procedures), Python, GitHook






Documentation

- Contains syntax, rules, code quality, pipeline, etc.
- Customized documentation for stakeholder.
- **RACI: Responsible, Accountable, Consulted, Informed.**

Task	Data Owner	Data Engineer	Data Consumer	Data Sensitivity
Create and maintain data dictionary	R	A	C	I
Ensure data quality and completeness	A	R	C	I
Build and maintain ETL pipeline	A	R	I	I
Set data naming and classification standards	A	R	I	C
Audit data	C	A	I	R

DIM_ASSETS	
asset_id 	integer
asset_name	varchar
length_or_area	decimal
asset_type	varchar
geometry_type	varchar
asset_status	varchar
is_active	boolean
install_date	timestamp
address	text
geometry	geometry



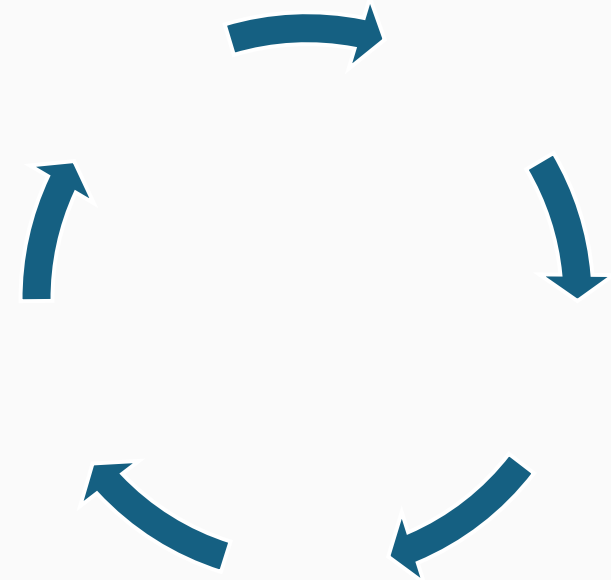
Maintainability

- Data Dictionary
- Asset Hierarchy Mapping
- Metadata Standards
- Governance Playbooks
- Data Flow Diagrams
- Versioning, centralized storage, stakeholder reviews

✓ *KISS – Keep It Stupidly Simple*
✓ *Modularity in model design*
✓ **Scalable Solutions, Cost-effective Approach (avoid rework)**

Asset Lifecycle

- **That which is not tracked, cannot be measured.**
- **High quality data capture, enables total cost of ownership.**
- Reactive Maintenance → Predictive Maintenance
- Iterative, collaborative, user-centric process.



References

Source	Referenced For
<u>OpenData Portal</u>	Structure of assets



Asset Hierarchy



**Asset Data
Model**



Data Dictionary



**Data Governance
Frameworks**



Questions

Thank You.

Questions?