

# Data Lineage in the IMX Orchestration of Federated Data Sources



**Auteur** Pano Maria

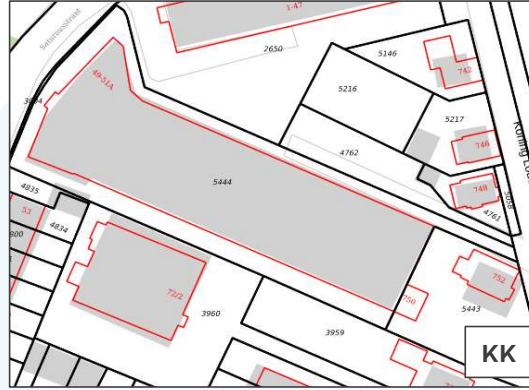
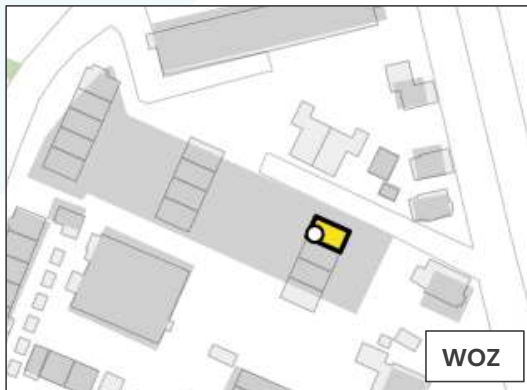
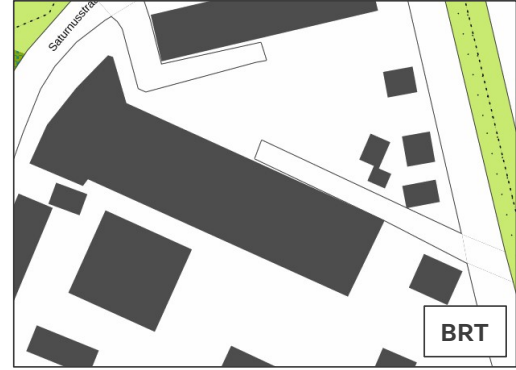
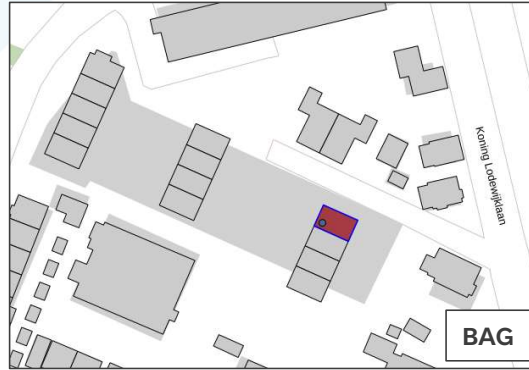
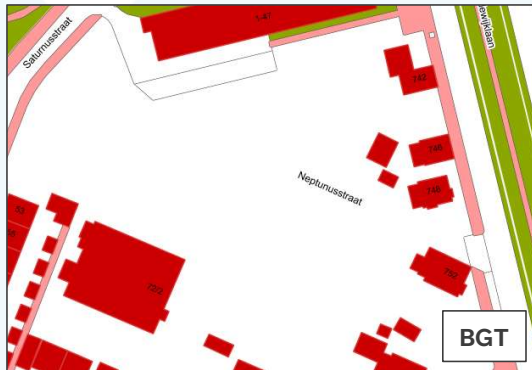
**Datum** Wednesday, September 6, 2023





Neptunusstraat, Apeldoorn, June 22 2019

# Geo base registries: Different views on reality



Neptunusstraat, Apeldoorn, June 22 2019

# From islands to integrated landscape





# How to achieve coherence?

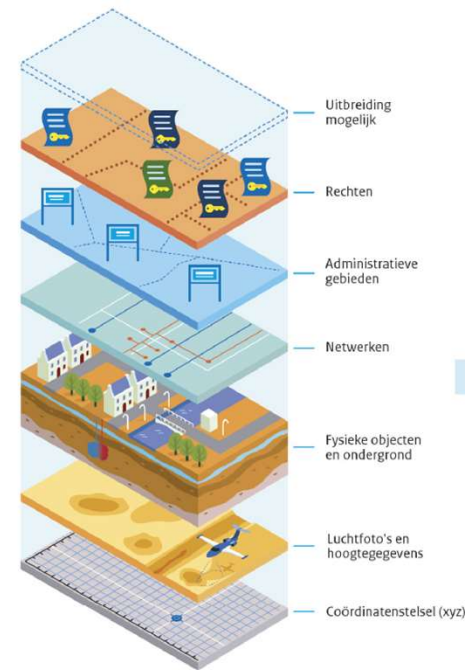
New registry with a coherent Geo view?

Emerging strategy on Federative data:

- data at the source
- stop copying data

Related programs:

- NGII
- Federatief datastelsel (IBDS)
- Zicht op Nederland (ZoN)
- Common Ground



Give me all information on this parcel

Energie  
Product API



Graafgebied  
Product API



Gebouw  
Product API



Adres  
Product API

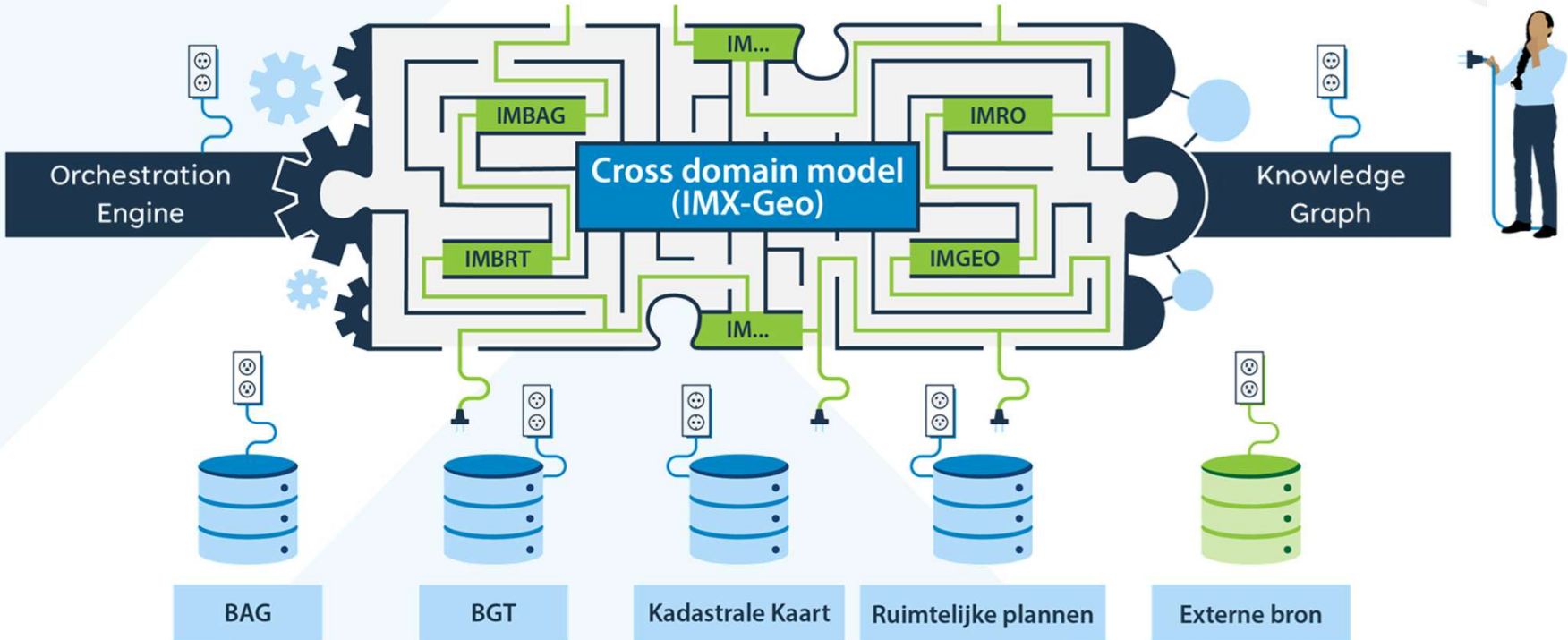


IMX

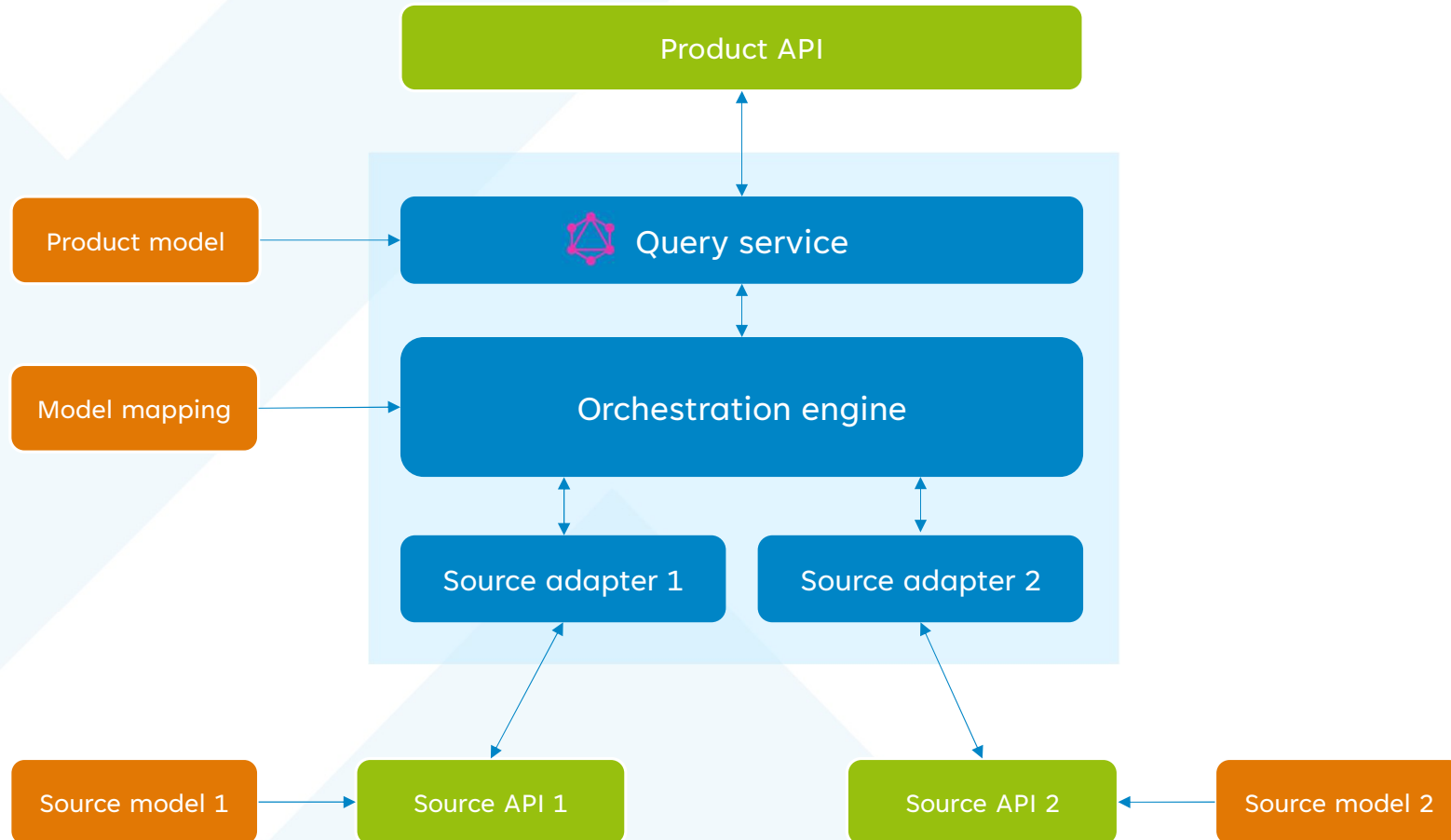


interactive  
instruments

Give me all homes with a purchase price lower than 400,000 euros in District A



# Orchestration engine



# Product model: IMX-Geo

- Cross-domain model
- Combines the Geo base registries without changing that data sources
- Introduces new relationships between underlying concepts

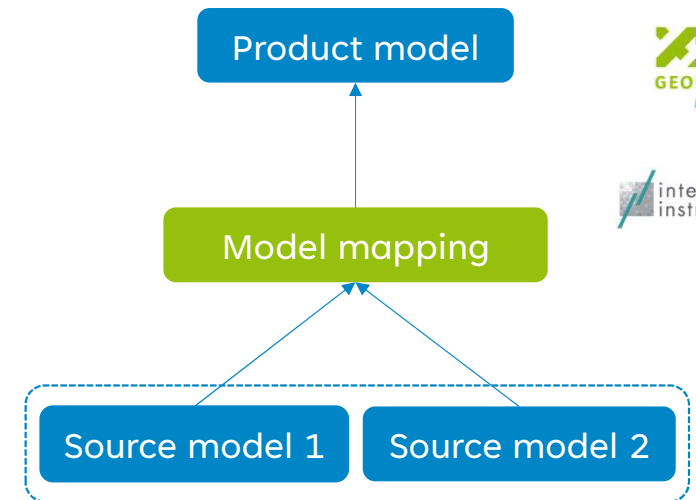




# IMX Model mapping

- Translates between product model and source model(s)
- Supports different modeling standards
  - MIM, OGC/ISO, RDFS/OWL/SHACL, ...
- Mapping and orchestration is stackable
  - A product model can become a source model
- Formal specification (emerging)
  - <https://github.com/Geonovum/IMX-ModelMapping/>

<https://github.com/Geonovum/WaU>



```
objectTypeMappings:  
  Adres:  
    sourceRoot: bag:Nummeraanduiding  
    propertyMappings:  
      identificatie:  
        pathMappings:  
          path: identificatie  
      huisnummer:  
        pathMappings:  
          path: huisnummer  
      huisnummertoevoeging:  
        pathMappings:  
          path: huisnummertoevoeging  
      huisletter:  
        pathMappings:  
          path: huisletter  
    postcode:  
      pathMappings:  
        path: postcode
```

# Orchestration Demo

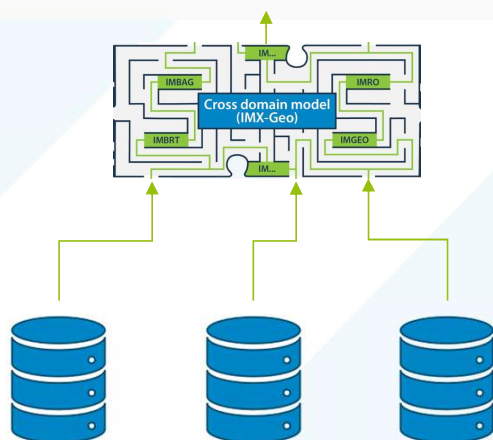
# Data Lineage

# Data lineage of data objects

```

{
  "omschrijving": "Laan van Westenenk 701, 7334DP Apeldoorn",
  "plaatsnaam": "Apeldoorn",
  "identificatie": "0200200000075716",
  "isHoofdadres": true,
  "huisnummer": 701,
  "postcode": "7334DP",
  "straatnaam": "Laan van Westenenk"
}

```



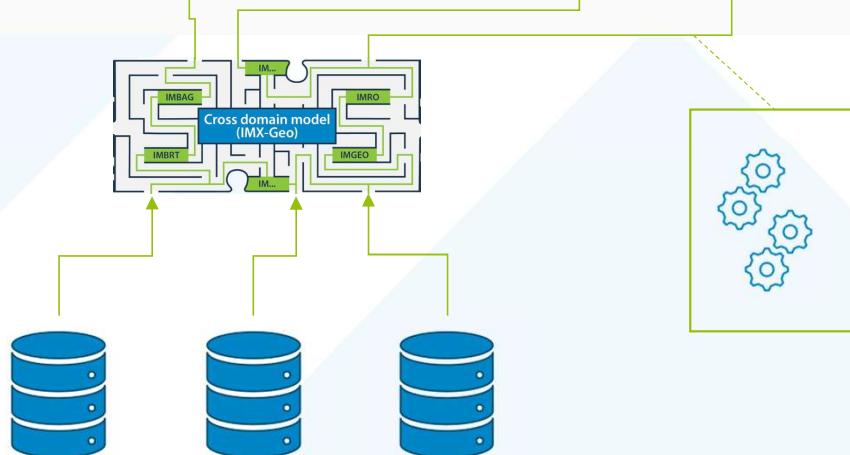
- Representation of the path along which data flows from origin to usage.
- Describes data transformations and processing.
- Represented by interlinked components such as data elements, business processes, and IT systems.

# Data lineage of data objects

```

{
  "omschrijving": "Laan van Westenenk 701, 7334DP Apeldoorn",
  "plaatsnaam": "Apeldoorn",
  "identificatie": "0200200000075716",
  "isHoofdadres": true,
  "huisnummer": 701,
  "postcode": "7334DP",
  "straatnaam": "Laan van Westenenk"
}

```



- Representation of the path along which data flows from origin to usage.
- Describes data transformations and processing.
- Represented by interlinked components such as data elements, business processes, and IT systems.



# Requirements for lineage of data elements

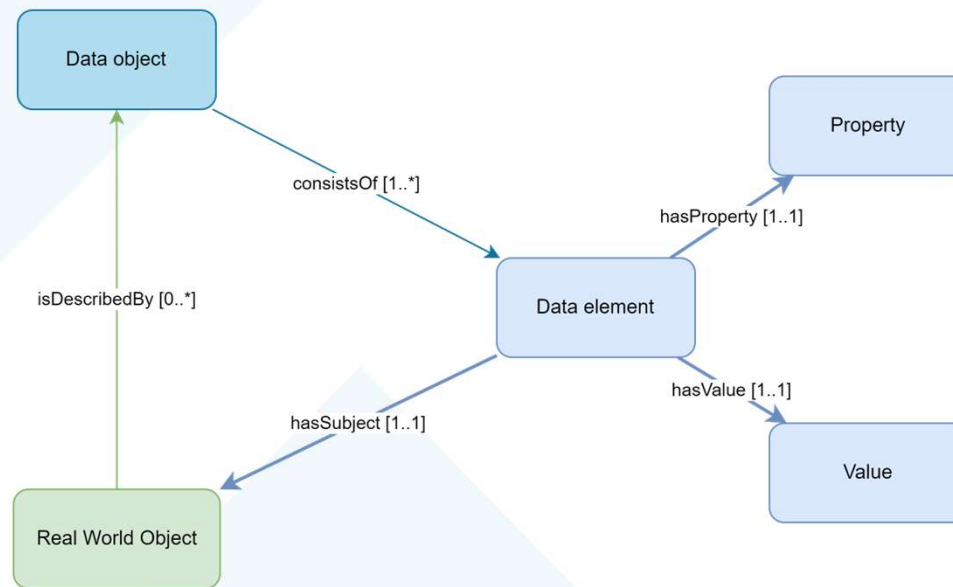
## Data lineage is metadata

- We need a way to model data elements
- Yet, keep our core data models simple
- We want to keep lineage separate from the data object itself
- We want data lineage to be publishable separately from the data object

For this we need:

1. A way to model data elements as objects
2. A data model for data lineage

# Modeling data elements as objects



# Example: data object as data elements

```
{  
  "identificatie": "0200200000075716",  
  "plaatsnaam": "Apeldoorn",  
  "isHoofdadres": true  
}
```



```
[  
  {  
    "subject": {  
      "identificatie": "0200200000075716"  
    },  
    "property": "identificatie",  
    "value": {  
      "stringValue": "0200200000075716"  
    }  
  },  
  {  
    "subject": {  
      "identificatie": "0200200000075716"  
    },  
    "property": "plaatsnaam",  
    "value": {  
      "stringValue": "Apeldoorn"  
    }  
  },  
  {  
    "subject": {  
      "identificatie": "0200200000075716"  
    },  
    "property": "isHoofdAdres",  
    "value": {  
      "booleanValue": true  
    }  
  }  
]
```

# Existing data lineage approaches

- Most data lineage approaches are tailored to data engineering in the data warehouse / data lake / data mesh space
- Job / run centric, focusing on columnar data

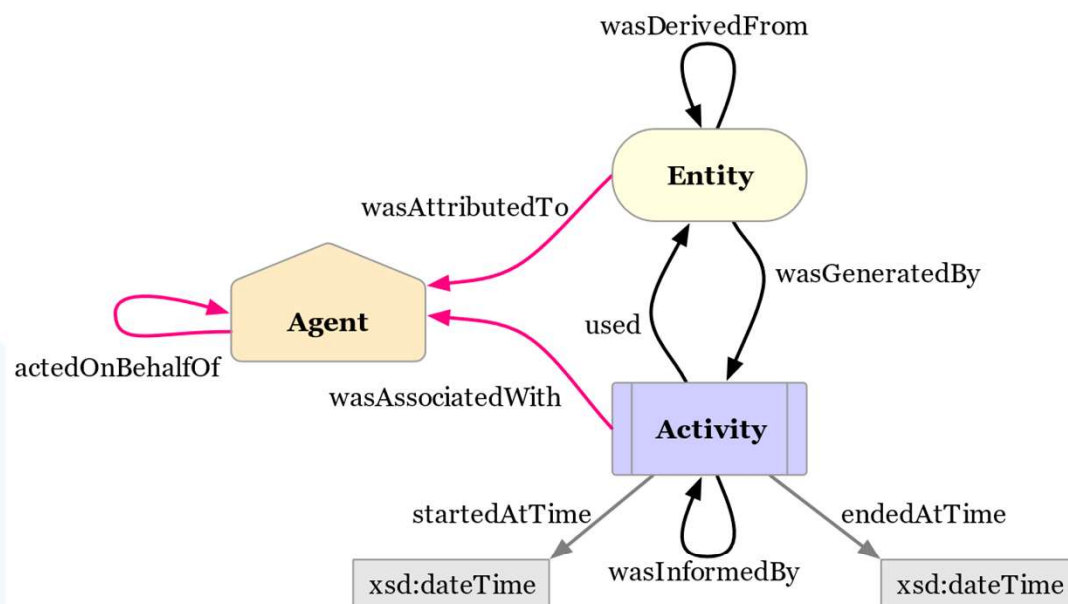


Open **Lineage**



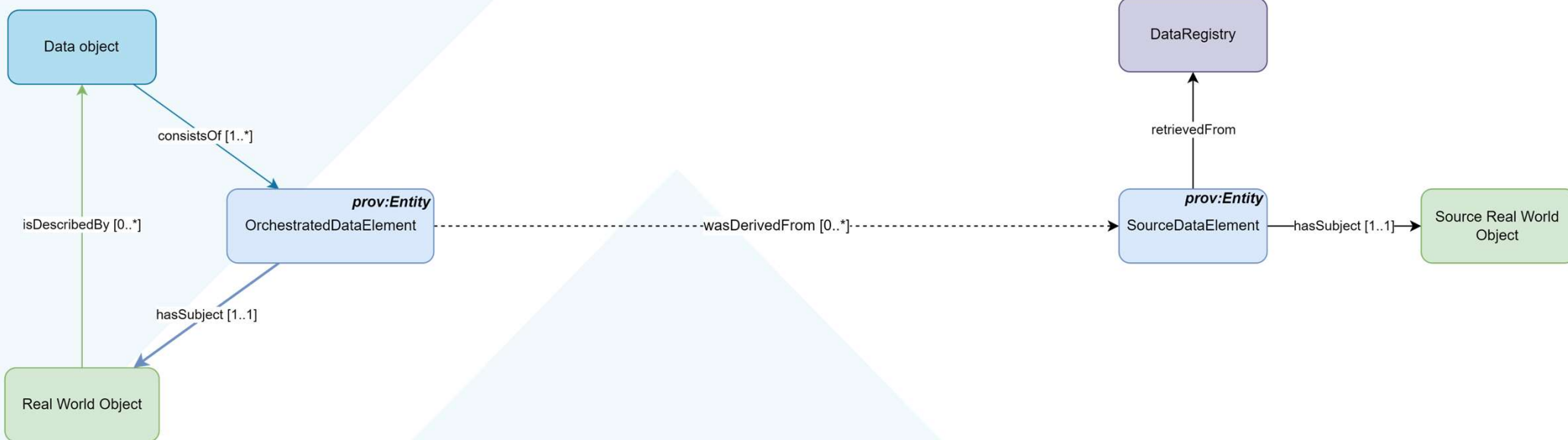
# Data lineage based on W3C PROV

- Simple extensible model
- No presumptions on architecture
- Must specialize to use case

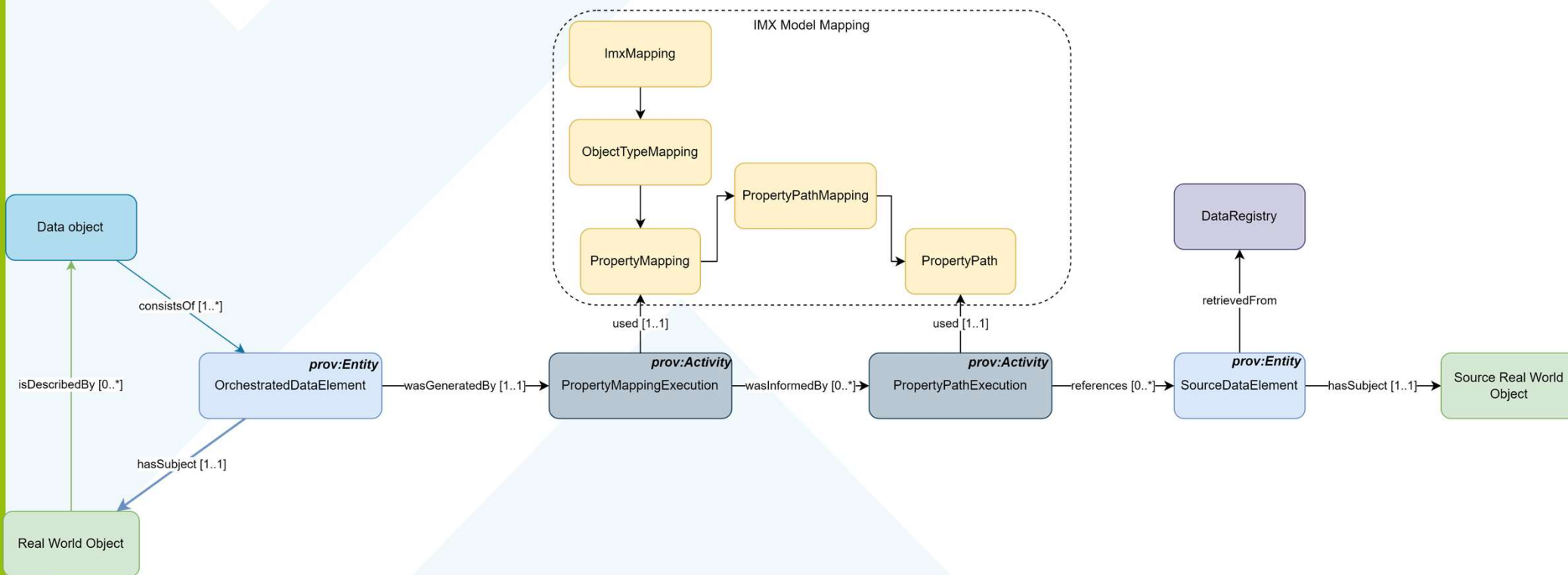




# Data lineage based on W3C PROV



# Data lineage based on W3C PROV



<https://geonovum.github.io/IMX-Metadata/>



INHOUDSOPGAVE

- Samenvatting
- Status van dit document
- 1. Introduction
  - 1.1 Target audience
  - 1.2 Introduction to document
  - 1.3 Working proces
- 2. Vocabulary
- 3. Metadata MIM extension
  - 3.1 Requirements
  - 3.2 Explanation and definition
  - 3.3 MIM-UML extension
  - 3.4 Implementation, encoding
- 4. Conformiteit
- 5. Lijst met figuren
- A. Index

## WaU - Metadata basic principles

Geonovum Handreiking  
Werkversie 19 juni 2023



**Laatste werkversie:**

<https://geonovum.github.io/wau-metadata>

**Redacteur:**

Paul Janssen (Geonovum)

**Auteur:**

Jesse Bakker (Kadaster)

**Doe mee:**

[GitHub Geonovum/NL-ReSpec-GN-template](#)

[Dien een melding in](#)

[Revisiehistorie](#)

[Pull requests](#)

Dit document is ook beschikbaar in dit niet-normatieve formaat: pdf



Dit document valt onder de volgende licentie:  
Creative Commons Attribution 4.0 International Public License

INHOUDSOPGAVE

- 1. Introduction
  - 1.1 Scope
  - 1.2 Working proces
- 2. Analysis
  - 2.1 Context
  - 2.2 Related or referenced standards
  - 2.3 Relevant documentation
- 3. Requirements and approach
  - 3.1 Requirements
  - 3.2 Approach
    - 3.2.1 Summary
    - 3.2.2 Limitations
- 4. Lineage - Information model
  - 4.1 Fundamentals
  - 4.2 Introduction to UML and objectcatalogue
- 5. UML and objectcatalogue
  - 5.1 Lineage - detail
  - 5.2 Objecttypen
    - 5.2.1 Georkestreerd/Gegeven
    - 5.2.2 Brongegeven
    - 5.2.3 BronRelatieGegeven
    - 5.2.4 Bronobject
    - 5.2.5 Bronregistratie

<https://geonovum.github.io/IMX-LineageModel>

## Lineage applied in WaU

Geonovum Informatiemodel  
Werkversie 16 juni 2023



**Deze versie:**

<https://github.com/geonovum/WaU-LIN>

**Laatst gepubliceerde versie:**

geen

**Laatste werkversie:**

<https://github.com/geonovum/WaU-LIN>

**Redacteurs:**

Paul Janssen, [Geonovum](#)

Naam Editor-n, [Geonovum](#)

**Auteurs:**

Pano Maria, [Geonovum](#)

Jesse Bakker, [Kadaster](#)

**Doe mee:**

[GitHub geonovum/WaU-LIN](#)

[Dien een melding in](#)

[Revisiehistorie](#)

[Pull requests](#)

**Rechtenbeleid:**



Creative Commons Attribution 4.0 International Public License  
(CC-BY)

# Publications

<https://github.com/Geonovum/WaU>

# Data Lineage Demo

## Future work

- Geonovum will organize an open consultation for IMX-Geo.
- IMX concept to be tested and further developed in <https://digilab.overheid.nl/>
- Talks with government agencies to develop the orchestration engine further. The orchestration engine will be open source.

Follow development: <https://github.com/imx-org/imx-orchestrate>



# Thanks!

**Pano Maria**

## Contact

**E-mail:** [p.maria@geonovum.nl](mailto:p.maria@geonovum.nl)  
[pano.maria@kadaster.nl](mailto:pano.maria@kadaster.nl)  
[pano@skemu.com](mailto:pano@skemu.com)

[www.geonovum.nl](http://www.geonovum.nl)

