ORGANISED BY

BDV BIG DATA VALUE ASSOCIATION

HOSTED BY

COSMOTE

DATAWEEK

JOIN.LEARN.SHARE.GET VALUE

Odyssey of AI:

Navigating the Data Seas

27 - 28 May / Athens

IN COLLABORATION WITH

ATC ATHENS TECHNOLOGY CENTER

ΕΠΙΣΕΥ ICCS

core|IC

DEMOKRITOS NATIONAL CENTRE FOR SCIENTIFIC RESEARCH

FORTH INSTITUTE OF CHEMICAL ENGINEERING SCIENCES

NEUROPUBLIC Information Systems & Technologies

data-week.eu
#DataWeek25

DATAWEEK

JOIN.LEARN.SHARE.GET VALUE

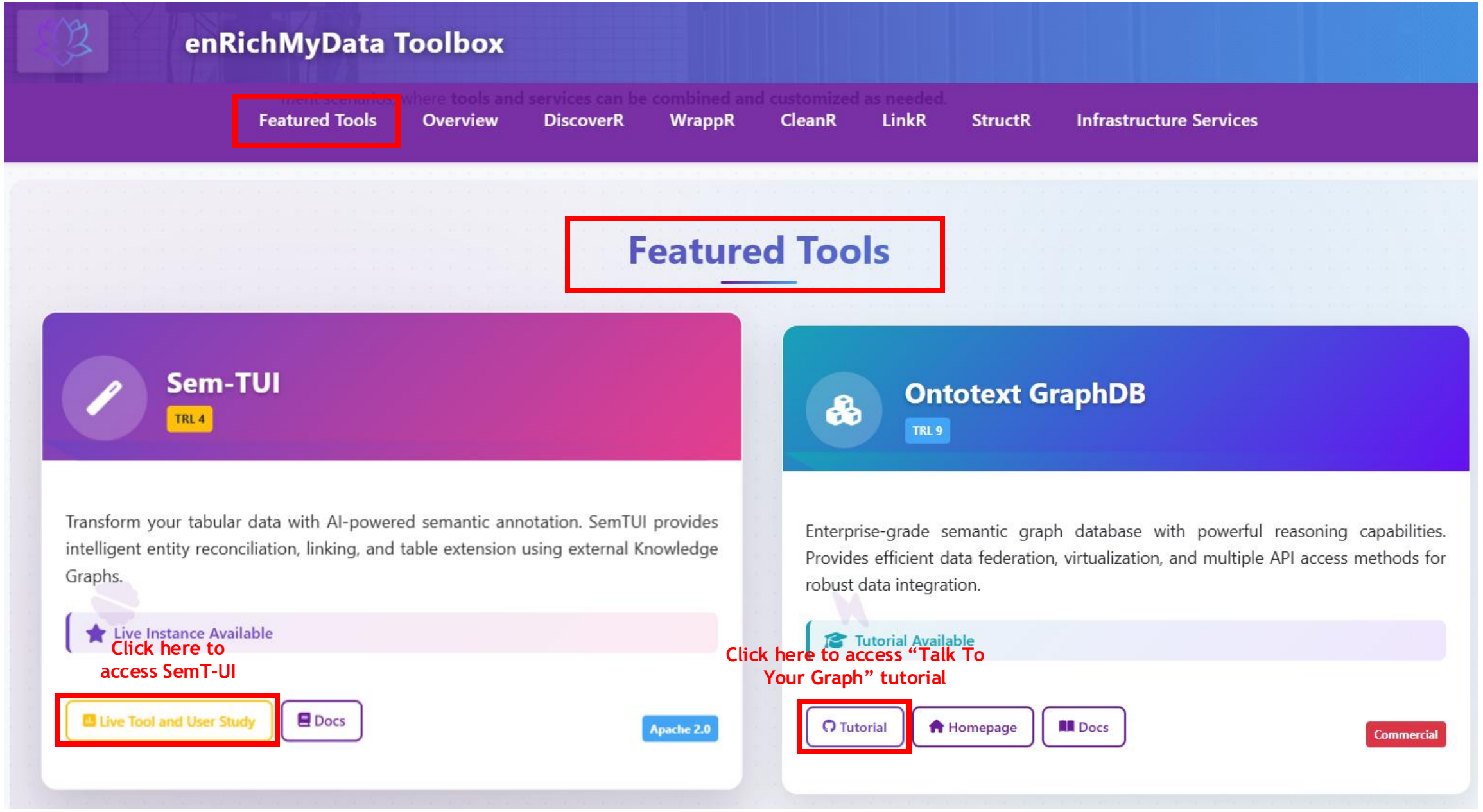
Data Enrichment for Industry: Unlocking the Potential of Reference Data

tinyurl.com/emdtoolbox

ORGANISED BY BDV BIG DATA VALUE ASSOCIATION

HOSTED BY COSMOTE

IN COLLABORATION WITH ATC, CORE|IC, FORTH, DEMOKRITOS, NEUROPUBLIC



DATAWEEK
JOIN.LEARN.SHARE.GET VALUE

DATAWEEK

JOIN.LEARN.SHARE.GET VALUE

Nikolay NIKOLOV
SINTEF

Nikola TULECHKI
Graphwise

Matteo PALMONARI
University of Milano-Bicocca

ORGANISED BY BDV BIG DATA VALUE ASSOCIATION

HOSTED BY COSMOTE

IN COLLABORATION WITH ATC, CORE|IC, FORTH, DEMOKRITOS, NEUROPUBLIC

Agenda

tinyurl.com/
emdtoolbox

1. Introduction to Data Enrichment (Nikolay Nikolov)
2. “Talk to Your Graph” (Nikola Tulechki)
3. Hands-on Enrichment Using SemT-UI (Matteo Palmonari)

tinyurl.com/
emdtoolbox

Introduction to Data Enrichment

Data Enrichment

tinyurl.com/

What is data enrichment?

Putting data in context, typically **add more data** to input data (structured / semi-structured / texts), improving data quality

Key
194
517
459
891
459

Compared to traditional data integration:

Data to enrich *with* (the reference datasets) are

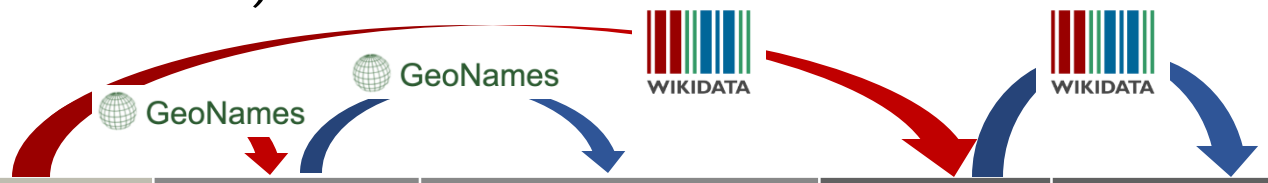
previously unknown to and **not under the control of** the **enricher**!

Challenges: ***Searching/finding***, ***understanding***, ***assessing***, ***accessing***

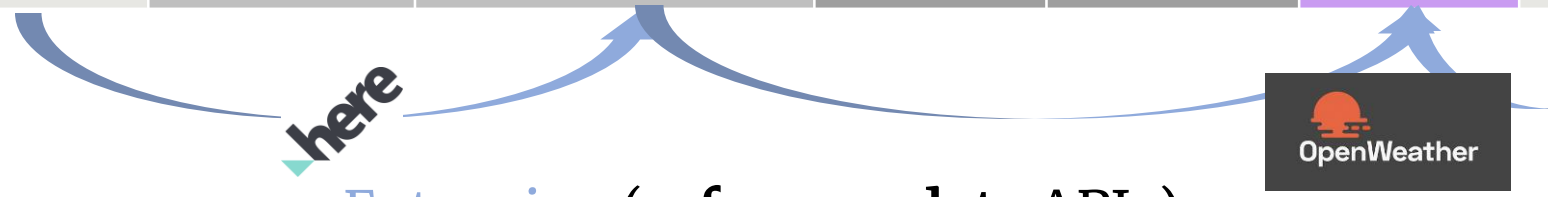
Data Enrichment: Augmentation

Reconciling with **reference data** (KGs or other data sources)

Extension (**reference data** from KGs)

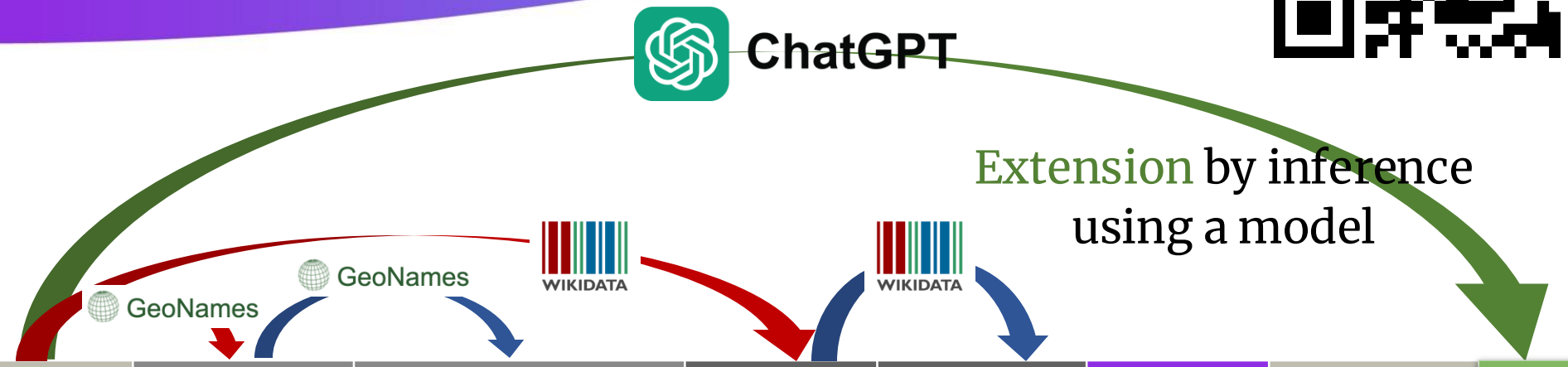


Keyword	#im	City	State	ID (Geonames)	Coordinates (Geonames)	ID (Wikidata)	Population (Wikidata)	Temp (OpenW)	Date
194906	64	Altenburg	Thuringia	2822542	[50.98763, 12.43684]	Q1205	2,122,335	18°	2017-03-11
517827	50	Inglostadt	Bavaria	2951839	[48.76508, 11.42372]	Q980	13,435,062	17°	2017-03-12
459143	42	Berlin	Berlin	2950157	[52.52437, 13.41053]	Q648102	3,782,202	17°	2017-03-12
891139	36	Munich	Bavaria	2951839	[48.13743, 11.57549]	Q980	13,435,062	19°	2017-03-11
459143	30	Nuremberg	Bavaria	2951839	[49.45421, 11.07752]	Q980	13,435,062	12°	2017-03-12

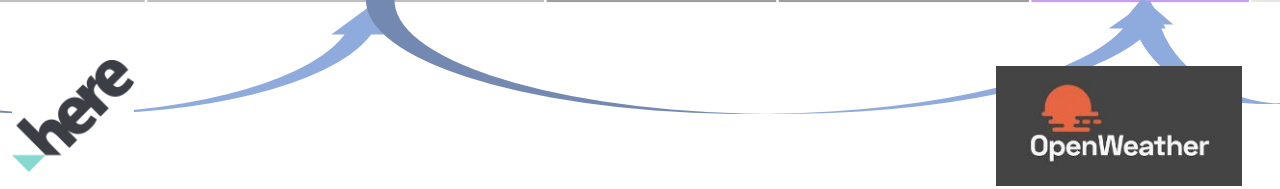


Extension (**reference data** APIs)

Data Enrichment: Augmentation

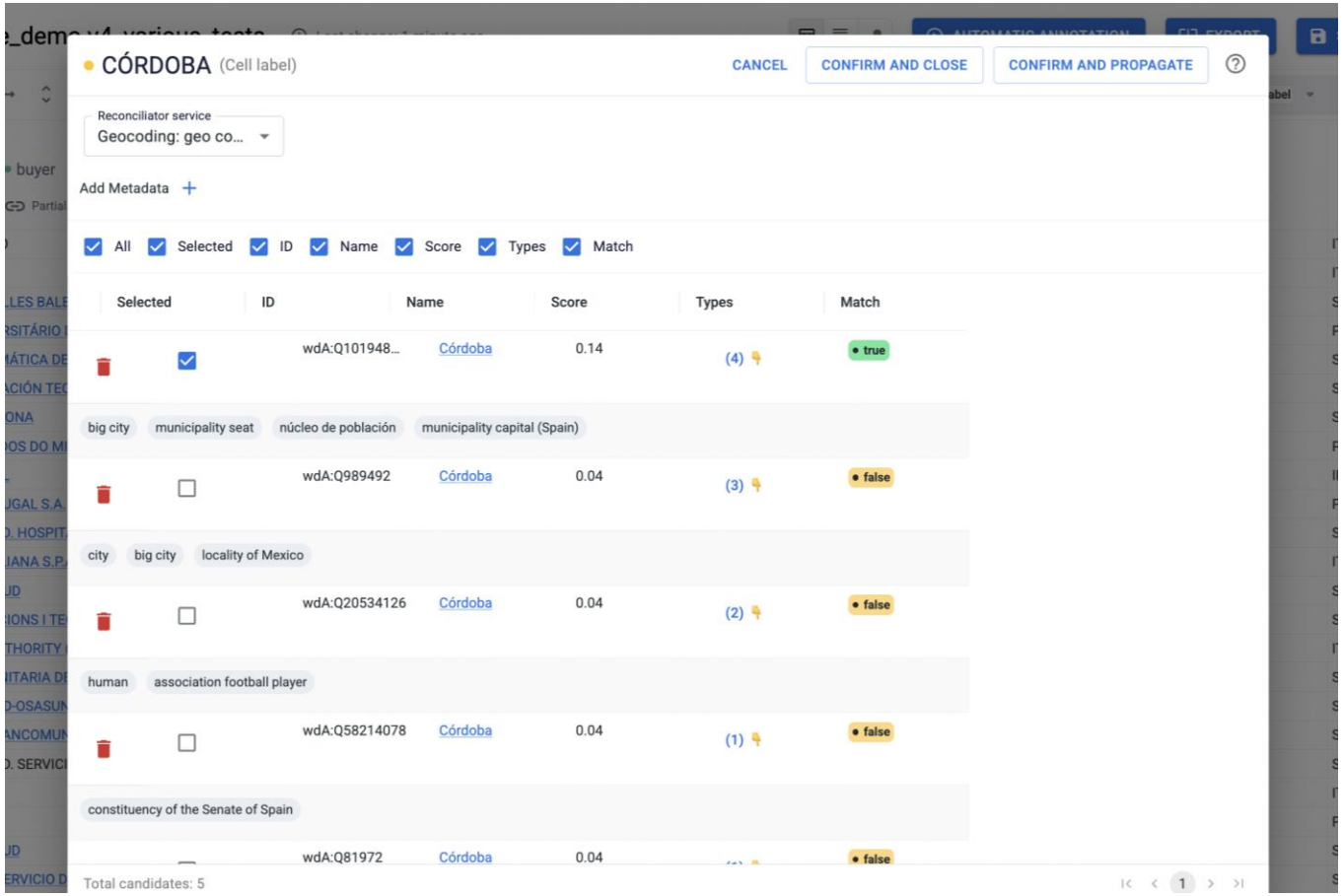


Keyword	#im	City	State	ID (Geonames)	Coordinates (Geonames)	ID (Wikidata)	Population (Wikidata)	Temp (OpenW)	Date	Area
194906	64	Altenburg	Thuringia	2822542	[50.98763, 12.43684]	Q1205	2,122,335	18°	2017-03-11	East
517827	50	Inglostadt	Bavaria	2951839	[48.76508, 11.42372]	Q980	13,435,062	17°	2017-03-12	West
459143	42	Berlin	Berlin	2950157	[52.52437, 13.41053]	Q648102	3,782,202	17°	2017-03-12	East
891139	36	Munich	Bavaria	2951839	[48.13743, 11.57549]	Q980	13,435,062	19°	2017-03-11	West
459143	30	Nuremberg	Bavaria	2951839	[49.45421, 11.07752]	Q980	13,435,062	12°	2017-03-12	East



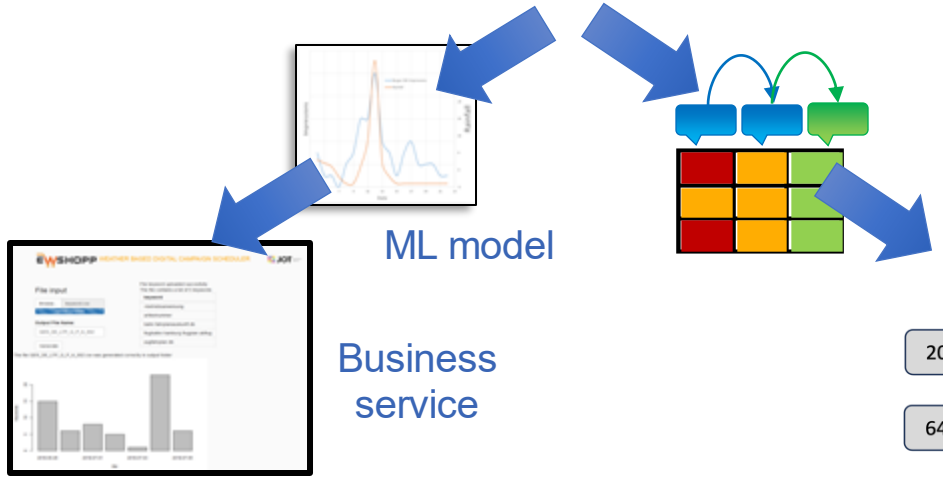
Why can Reconciliation be Hard?

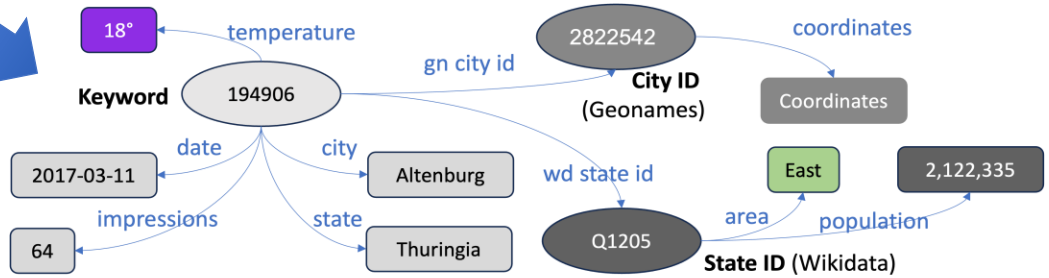
- Best candidate entities for “Còrdoba”
  - All have the same label
  - Different types
    - Big cities, province, football players
  - Similar entities with similar types
    - E.g., the province and the city giving the name to the province
  - Entities with same types but in different locations
    - Spain vs. Mexico



Data Enrichment: Transformation

Keyword	#im	City	State	ID (Geonames)	Coordinates (Geonames)	ID (Wikidata)	Population (Wikidata)	Temp (OpenW)	Date	Area
194906	64	Altenburg	Thuringia	2822542	[50.98763, 12.43684]	Q1205	2,122,335	18°	2017-03-11	East
517827	50	Inglostadt	Bavaria	2951839	[48.76508, 11.42372]	Q980	13,435,062	17°	2017-03-12	West
459143	42	Berlin	Berlin	2950157	[52.52437, 13.41053]	Q648102	3,782,202	17°	2017-03-12	East
891139	36	Munich	Bavaria	2951839	[48.13743, 11.57549]	Q980	13,435,062	19°	2017-03-11	West
459143	30	Nuremberg	Bavaria	2951839	[49.45421, 11.07752]	Q980	13,435,062	12°	2017-03-12	East





enRichMyData

Enabling Data Enrichment Pipelines for AI-driven Business Products and Services

SINTEF

PHILIPS

BOSCH

JOT INTERNET MEDIA

ES GROUP

expert.ai

ontotext

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA

Jožef Stefan Institute
Ljubljana, Slovenia

University of Copenhagen

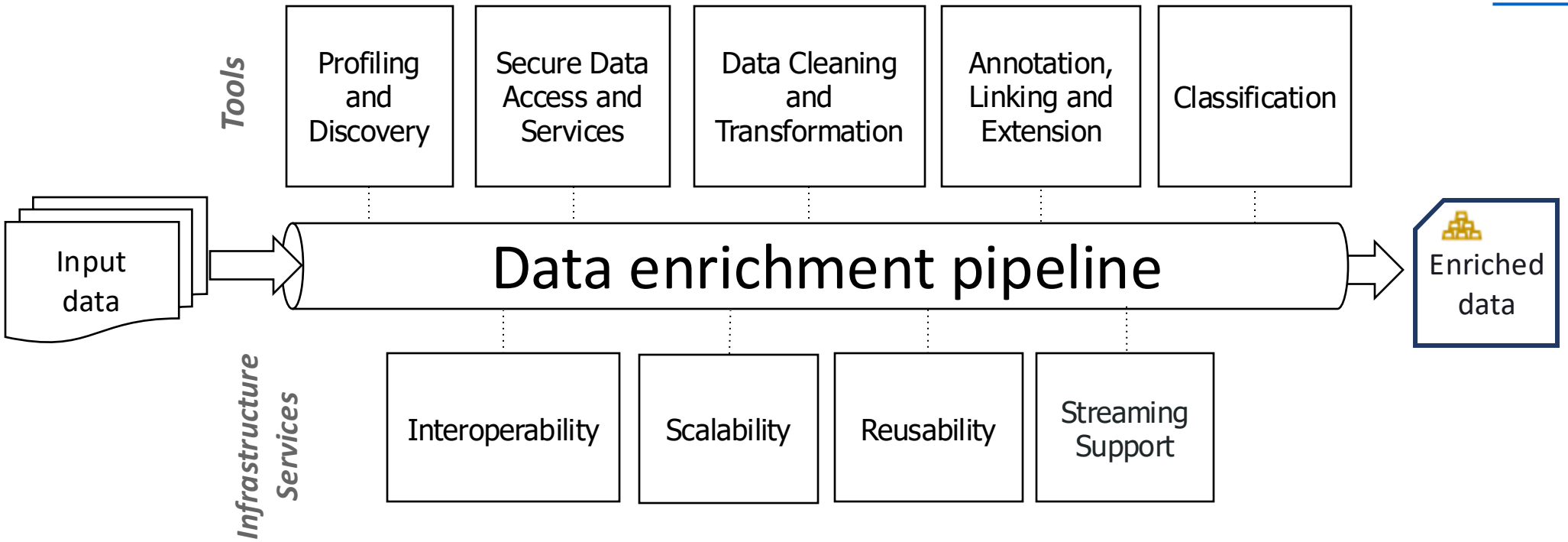
GATE

BGRIMM TECHNOLOGY GROUP

Spend Network.

enRichMyData Toolbox

tinyurl.com/
emdtoolbox



Presented by
Nikola Tulechki

Talk To Your Graph

How to Build a LLM Chatbot over KG

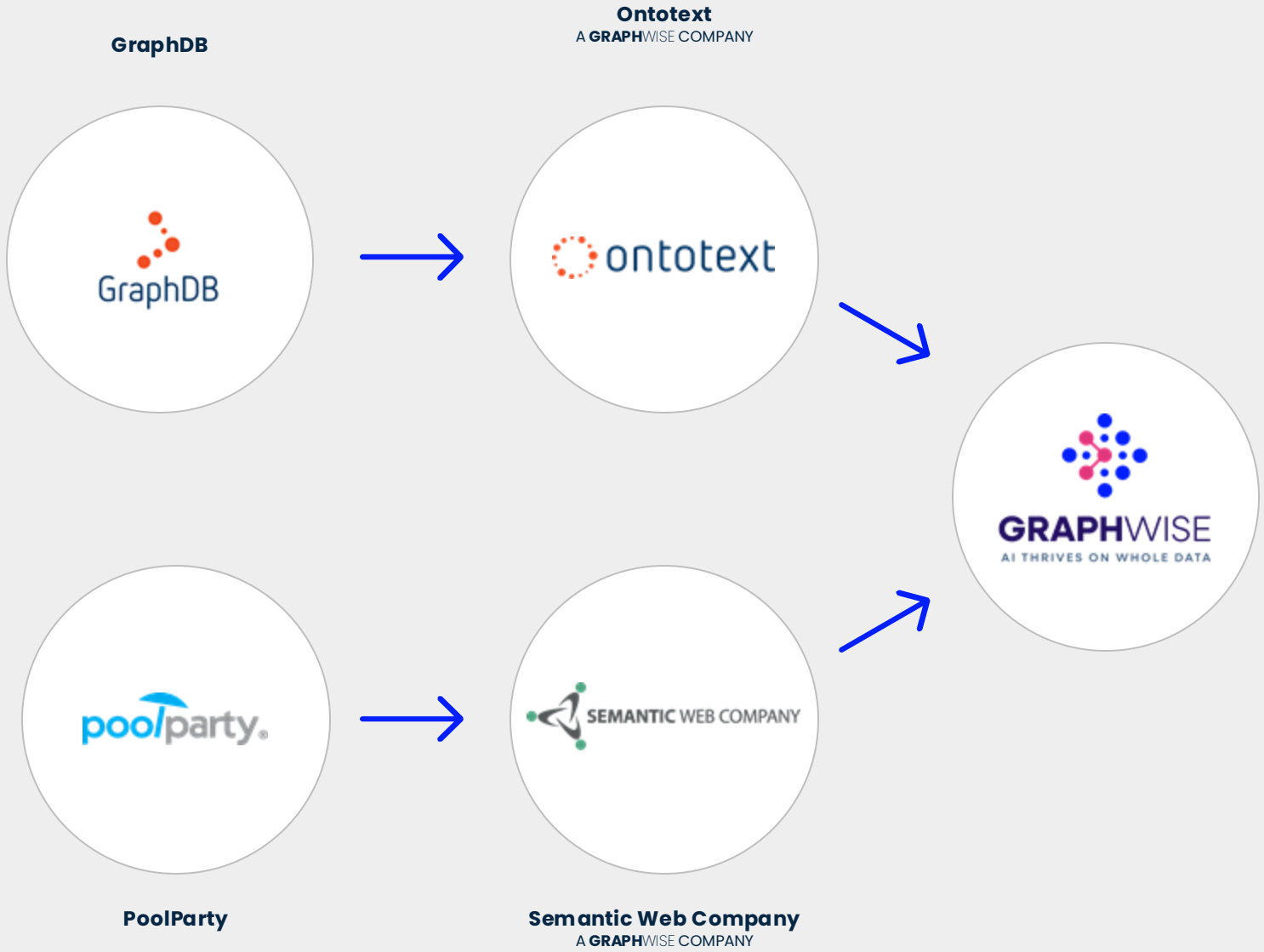
Athens Data Week - EC - EnrichMyData

Ontotext & SWC: the Driving Forces of Graphwise

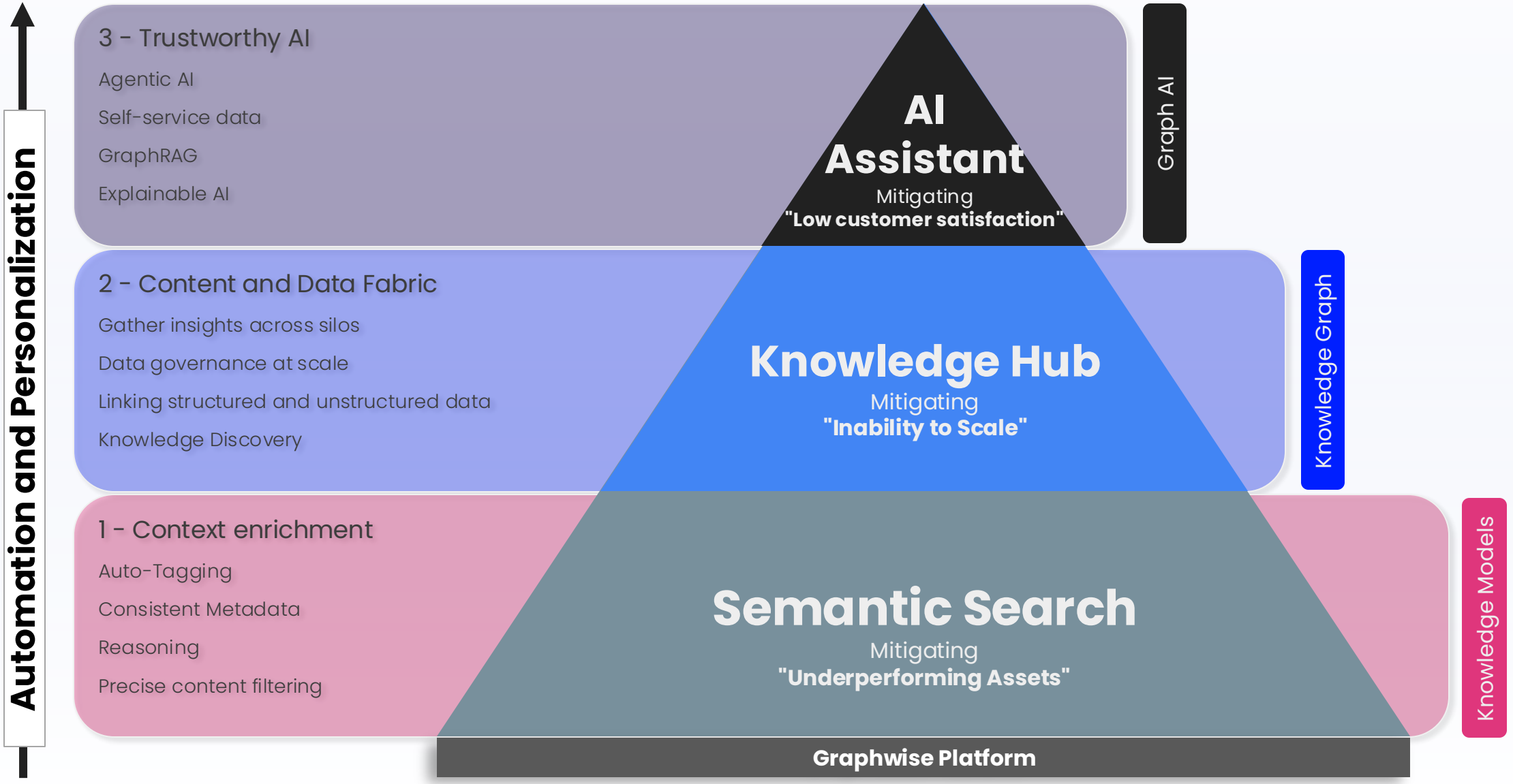
ontotext

One stop shop for Graph AI

Graphwise has the platform and tools, the methodologies, and the people to deliver world-class end-to-end solutions to enterprises.



Graphwise Value Pyramid



Graphwise tools in EnrichMyData toolbox

- **GraphDB** - triplestore semantic database
  - W3C RDF/SPARQL compliant
  - Data Virtualization
  - GraphQL interface
  - Full Text Search
  - *Interfaces with LLM for data interaction*
- **Ontotext Refine** - RDF data transformation tool based on OpenRefine
  - GUI and SPARQL declarative mappings
  - Command Line Interface
- **Ontotext Reconcile** - Reconciler compatible with Reconciliation API protocol

The Use Case and Data

Open Opportunities

- **Government Contracting Data**
  - Tenders / Buyers / Goods and services
  - Nightly crawl from hundreds of sources
- **Data Enrichment**
  - Matching to Wikidata (reconciliation Buyer name (LamAPI)
  - Matching to Geonames (reconciliation Ontotext Reconcile)
  - Matching to CPV thesaurus (via code ID)
- Data transformation to RDF Knowledge Graph
  - OntoRefine - JSON->RDF

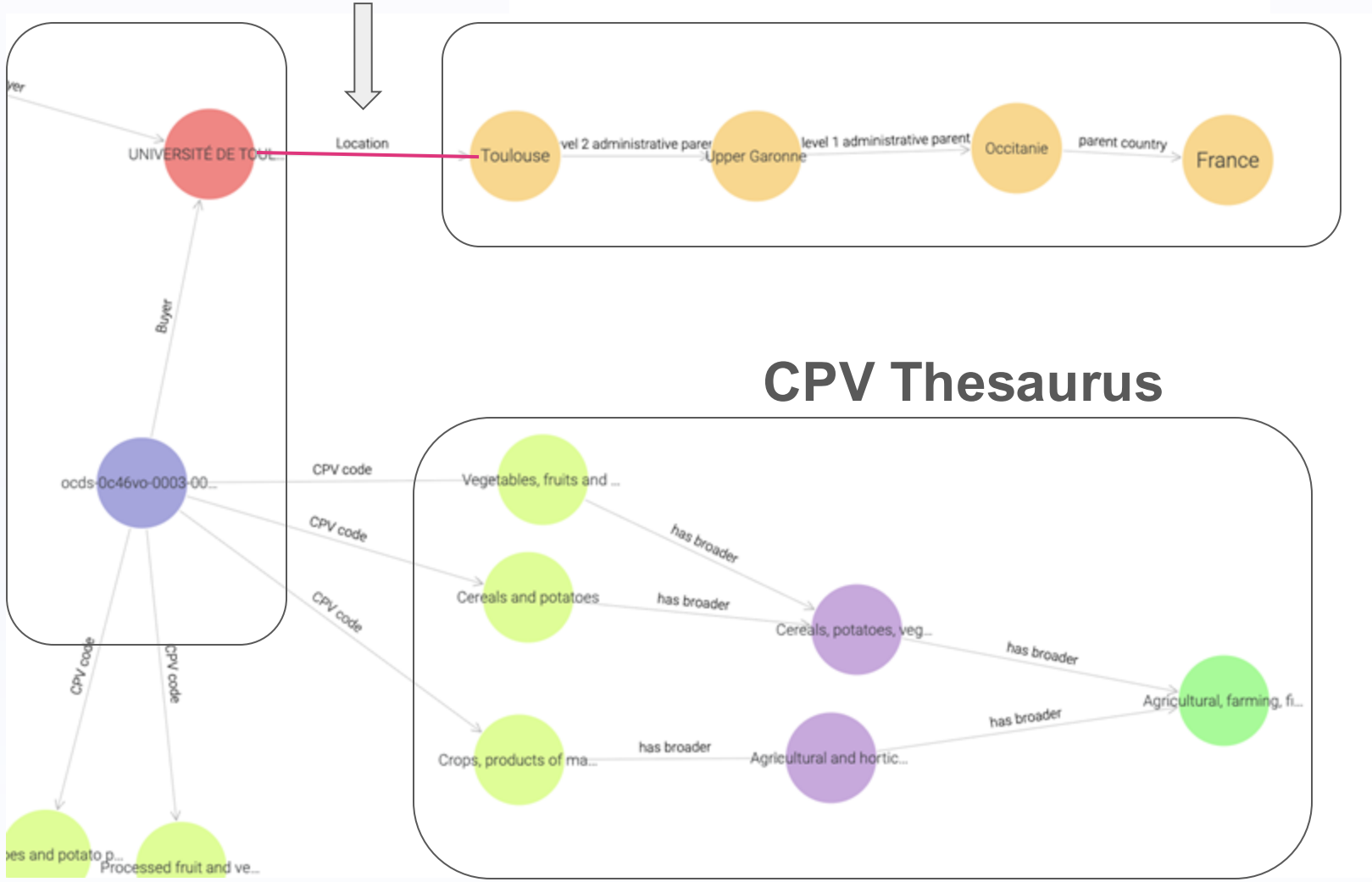
The Knowledge Graph

Ontotext Refine RDF Transformation


```
{
    "buyer_address_country_name": "FRANCE",
    "award_amount_sum": 38000,
    "buyer_name": "UNIVERSIT\u00c9 DE TOULOUSE",
    "buyer_address_locality": "Toulouse",
    "award_duration_in_days_first": null,
    "release_date": "2025-04-08T02:00:00+00:00",
    "tag_status": "closed",
    "tender_url": "https://ted.europa.eu/udl?uri=TED:NOTI(
    "date_updated": "2025-04-15T11:34:08.532871",
    "tender_id": "00239680-2025_td",
    "cpv_codes": [
      "15300000",
      "03220000",
      "15330000",
      "03210000",
      "15310000",
      "03110000"
    ],
     "ocid": "ocds-0c46vo-0003-00239680-2025",
    "tender_description": "Le pr\u00e9sent accord-cadre a
}
```


Reconcile

GeoNames



ONTOLOGY

Talk to Your Graph 2.0

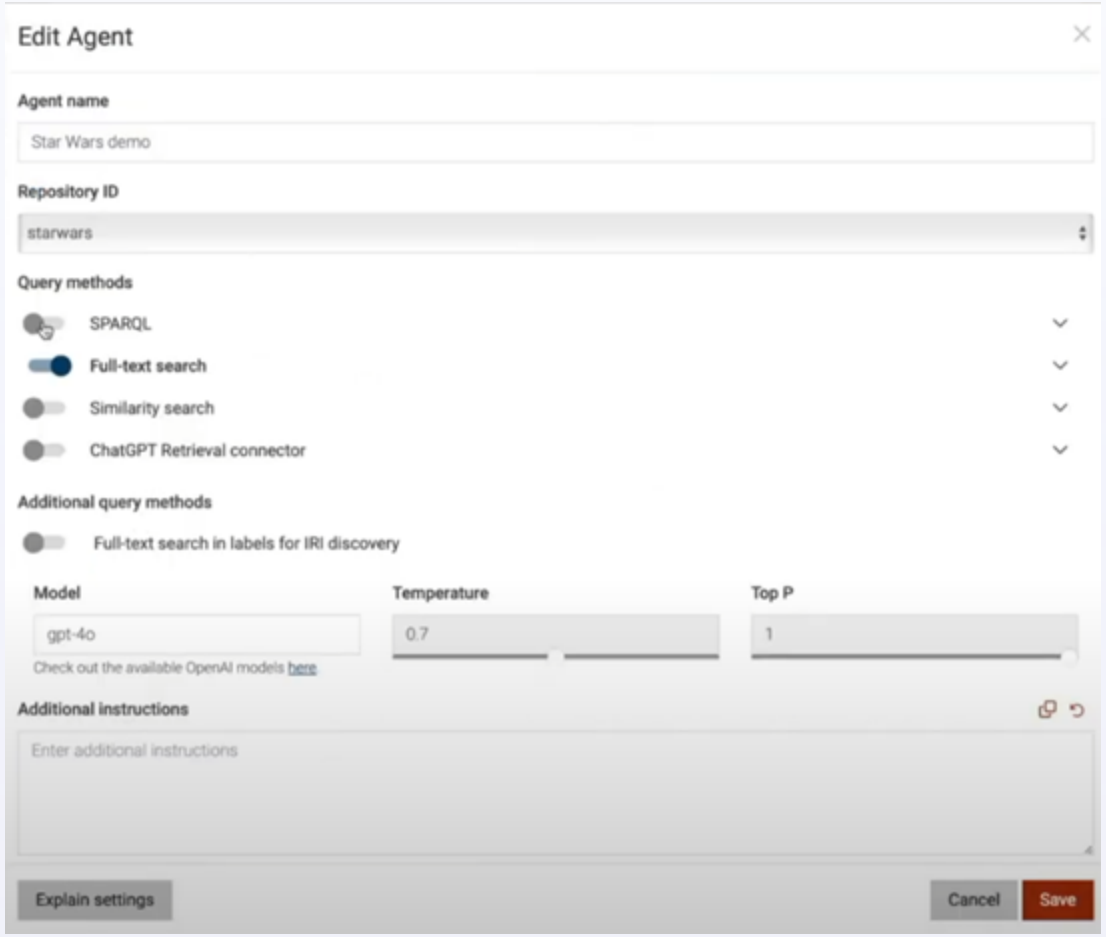
GRAPHWISE
AI THRIVES ON WHOLE DATA

Conversational AI for Knowledge Graphs: Enable natural language queries with SPARQL precision and versatile retrieval methods.

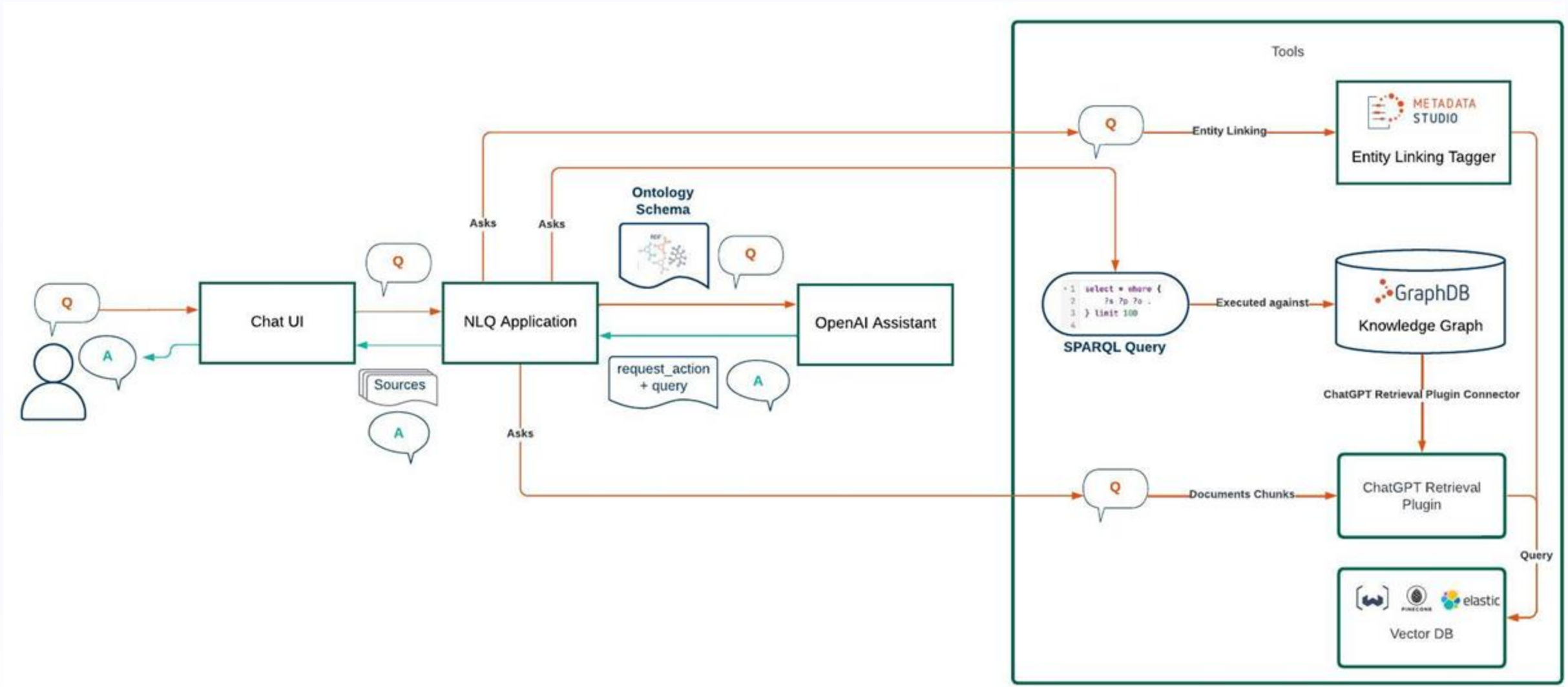
Multi-Method Querying: Combine SPARQL, vector similarity, and full-text search for dynamic data retrieval.

Customizable, Intelligent Agents: Combine different retrieval methods to deliver conversational insights and improve explainability

Empowering Users of All Skill Levels: Instant insights without technical expertise, reducing reliance on data teams.



Talk To Your Graph Basics



Demo Time

enRichMyData

DatAI Lab
Department of Informatics, Systems and Communication
University of Milan - Bicocca

SemT-UI: Demonstration and Usability Survey

Flavio De Paoli, Matteo Palmonari

*This work presented in this presentation has received funding from the European Union's Horizon 2020 research and innovation program under grant agreements No 732590 - **EW-Shopp** - and No 732003 – **euBusinessGraph** - and from the European Union's Horizon Europe research and innovation program under grant agreements No 101070284 - **enRichMyData**.*

DATAWEEK

JOIN.LEARN.SHARE.GET VALUE

Data Enrichment for Industry: Unlocking the Potential of Reference Data

tinyurl.com/emdtoolbox

ORGANISED BY BDV BIG DATA VALUE ASSOCIATION

HOSTED BY COSMOTE

IN COLLABORATION WITH ATC, CORE|IC, FORTH, DEMOKRITOS, NEUROPUBLIC

… part of a broader framework to support data enrichment

SemT-X framework

SemT-X at a Glance

Supported tasks

- Data enrichment
  - Reconciliation / linking / annotation
  - Data extension
- Two main application scenarios
  - Augmentation
  - Annotation and transformation

Data enrichment *simplified*

- The backend
  - Several services (in-house and third-party)
  - Open to integration of new services
- The frontends
  - **SemT-UI**: GUI for data discovery
  - SemT-Py: API for Jupyter notebook

SemT-X at a Glance

Supported tasks

- Data enrichment
  - Reconciliation / linking / annotation
  - Data extension
- Two main application scenarios
  - Augmentation
  - Annotation and transformation

Data enrichment *simplified*

- The backend
  - Several services (in-house and third-party)
  - Open to integration of new services
- The frontends
  - **SemT-UI**: GUI for data discovery
  - SemT-Py: API for Jupyter notebook

Larger data

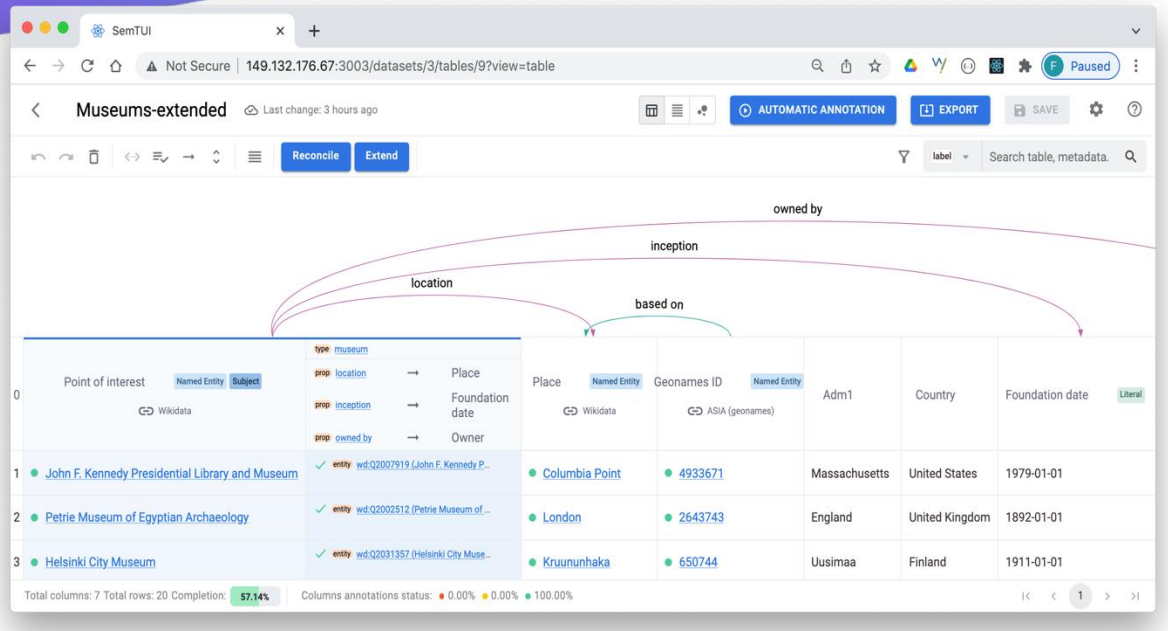
… even larger data via deployment on a distributed infrastructure like **Tao**

SemT-UI: Interactive Enrichment of Tabular Data

UI accessing external services

- Complete semantic table annotations
- Reconciliation/linking services
  - Linking: Wikidata (Alligator)
  - Linking: Wikidata (OpenRefine)
  - Geocoding: coordinates (GeoNames)
  - Geocoding: coordinates (HERE)
  - …

  … based on OpenRefine (W3C Specs)
- Extension services
  - Wikidata
  - Weather extension (OpenMeteo)
  - Distances and routes (HERE)
  - …



Support to *Linking – Revision – Extension* of tabular data

- Graphical view & revision of annotations
  - Global and specific annotation rendering
  - Single cell editing / annotation revision
  - Column annotation revision

Join SemT-UI usability test!
Help us improve the tool ☺

Testing + Feedback Collection

- Online questionnaire
- You will find
  - Link to a short video
    - main functionalities of the tool
  - Tasks to complete
    - Automatic reconciliation
    - Revision of links
    - Extension
  - Questions
- Completion time: ~20+ minutes
  - Additional optional tasks
- This demo
  - Walk through main functionalities, similar as in the video

tinyurl.com/semtuisurvey

Covered in the Demonstration

- Data sample from SpendNetwork
  - Buyers in public tenders with come contextual information
- Reconciliation and extension with Wikidata
  - Wikidata: a large open-source knowledge graph with 117M+ data items
- We will focus on
  - Reconciliation of buyer's names (public organizations publishing tenders) using an in-house algorithm
  - Revision of links to Wikidata by type inspection and manual editing
  - Extension using values from Wikidata properties
- Reminder
  - SemT-UI provide access to several services for data extension, not just to open-source knowledge graphs, e.g., Weather, Geocoding, etc.

Some Important Remarks

Account and password

- An account is needed to log into SemT-UI and complete the questionnaire
- How to get an account
  - Follow the instructions in the usability survey

Processing time and potential bugs

- Reconciliation algorithms may take some time, pls wait
- Some third-party services may encounter problems
- We are still testing, the application may have some bugs

Link to enRichMyData toolbox and SemT-UI

tinyurl.com/emdtoolbox

Direct link to usability survey

tinyurl.com/semtuisurvey