Challenges in Data Management for Evolving Knowledge Graphs

David Chaves-Fraga
CiTIUS@University of Santiago de Compostela (Spain)
david.chaves@usc.es

with the contributions from: Anastasia Dimou, Oscar Corcho,
Ana Iglesias-Molina, Xuemin Duan, Maria-Esther Vidal, and more...

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

CAMPUS VIDA CAMPUS DE EXCELENCIA INTERNACIONAL

CiTIUS Singular Research Center on **Intelligent technologies**

Who am I?

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Assistant Professor (Sept 2023-now)

PhD & Postdoc (2015-2023)

Ontology Engineering Group

Senior Researcher (Sept 2023-now)

Co-chair (2019-now)

KGC Workshop

W3C

DTAI Declarative Languages & Artificial Intelligence

KU LEUVEN

Postdoc & Research Collaborator (2021-now)

Research Interest

- Mapping Languages
- Data Integration at scale
- Benchmarking KG Construction
- Data Management for KG/SW

Is Data Management "sexy"?

Whether data management is considered "sexy" is a matter of perspective and can vary depending on the context and individual preferences.

Data management is a critical and foundational aspect of any organization's operations. It involves the collection, storage, organization, and analysis of data to make informed decisions and improve business processes. While some people may find the field of data management to be intellectually stimulating and crucial to the success of a business or organization, others may not see it as glamorous or exciting compared to more high-profile tech fields like artificial intelligence, cybersecurity, or data science.

In recent years, there has been a growing appreciation for the importance of data and information management, driven by the increasing volume and complexity of data that organizations deal with. Data governance, data privacy, and data quality are becoming essential components of many businesses, and skilled data professionals are in demand.

So, whether data management is "sexy" or not may depend on your perspective and the specific aspects of data management that interest you. It might not have the same appeal as some other cutting-edge technologies, but it plays a vital role in modern organizations and can be intellectually rewarding for those who are passionate about it.

be more clear with your answer

Data management is not traditionally considered a "sexy" or glamorous field in the same way that some emerging technologies or high-profile industries might be. It doesn't always receive the same level of public attention or excitement as, for example, artificial intelligence, virtual reality, or space exploration. However, this does not diminish its importance. Data management is a foundational and critical aspect of modern organizations, and skilled data professionals are in high demand.

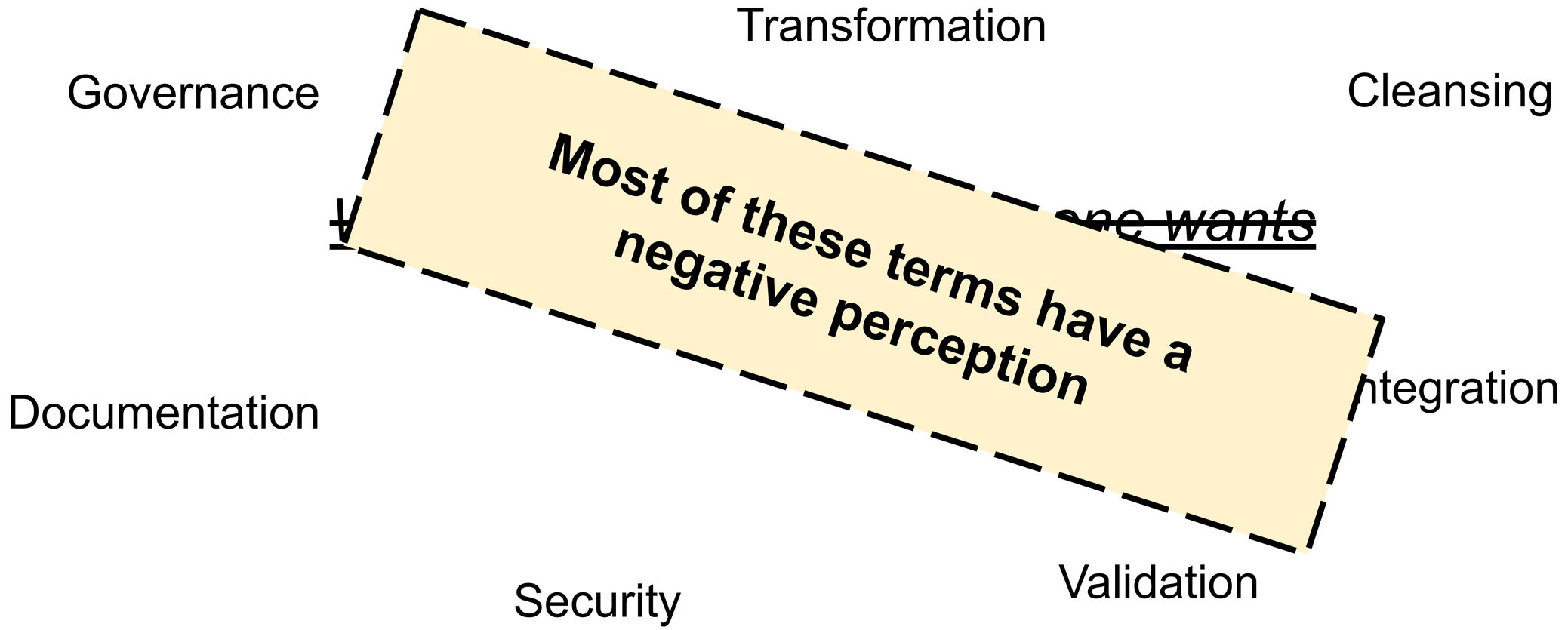
So, in popular culture or mainstream perception, data management may not be considered "sexy," but it is undeniably valuable and essential for the effective functioning of businesses and organizations in today's data-driven world.

So... what is data management?

*What we need to do but no one wants**

* from our pop culture

So… what is data management?



Real-world projects...

2020

Construct:
- an ontology network for DevOPS,
- a knowledge graph conformance with the ontology and,
- a QA system on top of the KG

2021

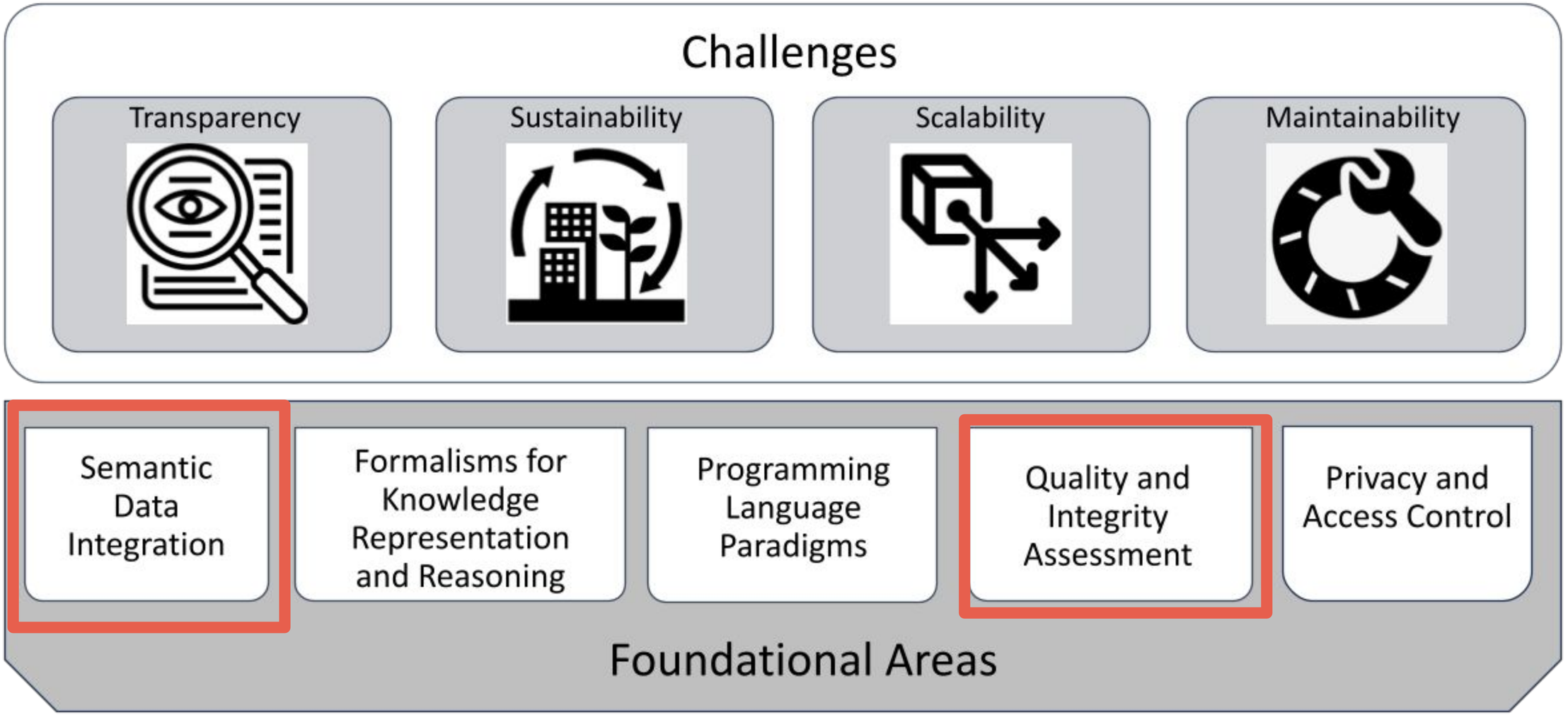
Construct:
- a knowledge graph conformance with a external/inmature ontology
- use an Oracle RDBMS with almost 2000 tables
- provide a mechanism to maintain the mappings

2022-now

EU Public Procurement Data Space

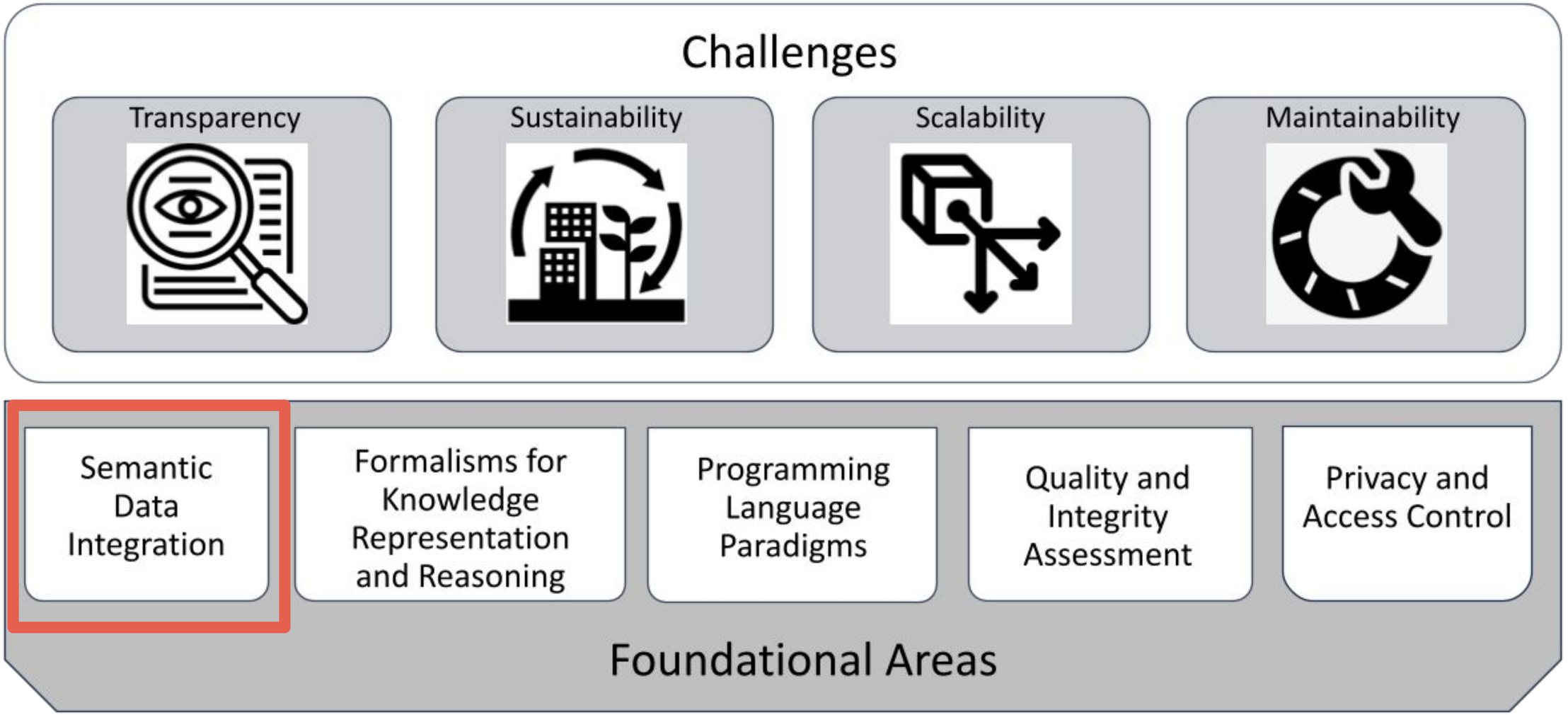
Construct:
- a decentralized KG-driven data ecosystem
- map all EU PP data w.r.t. ePo ontology (which evolves)
- everything needs to be on production and publicly available

Let’s make data management “sexy”…



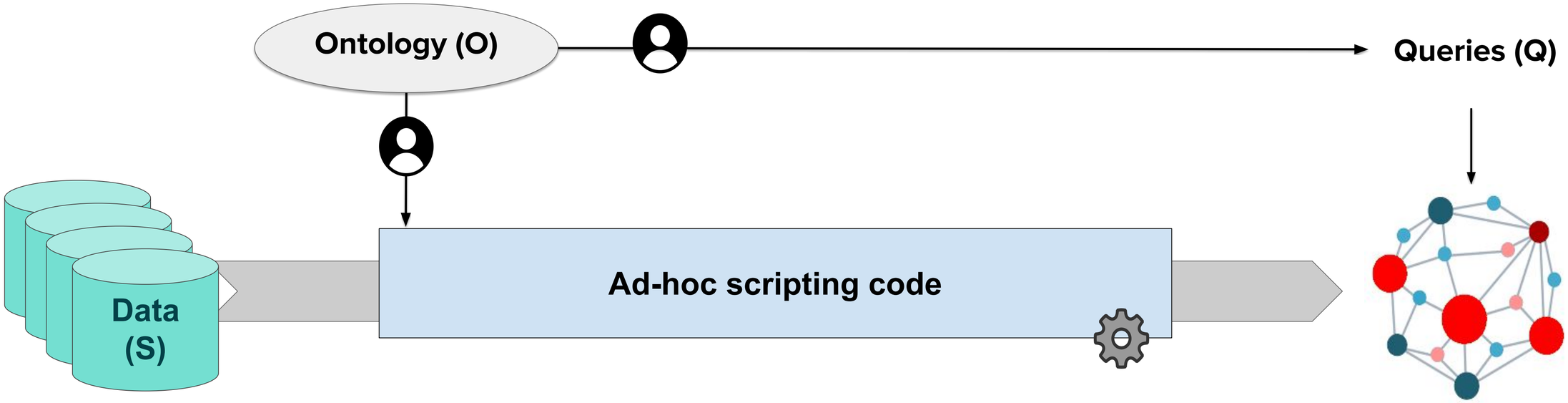
Chaves-Fraga, D., Corcho, O., Dimou, A., & Vidal ME. (2023). Challenges on Data Management for Knowledge Graphs (Under Review)

Let’s make data management “sexy”…



Chaves-Fraga, D., Corcho, O., Dimou, A., & Vidal ME. (2023). Challenges on Data Management for Knowledge Graphs (Under Review)

The construction of KG (v1.0)



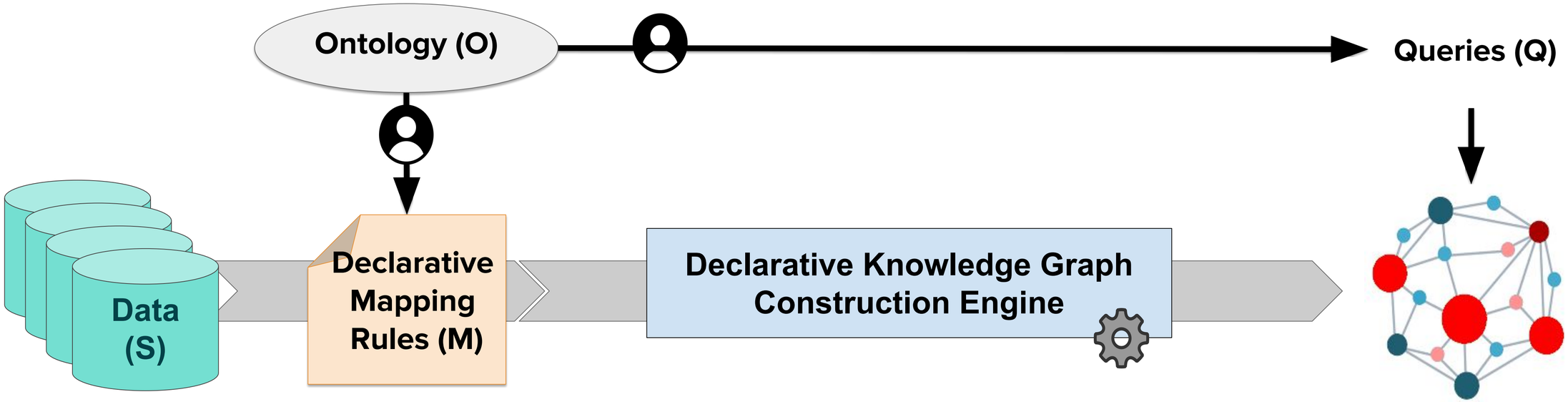
Sustainable

Scalable

Maintainable

Transparent

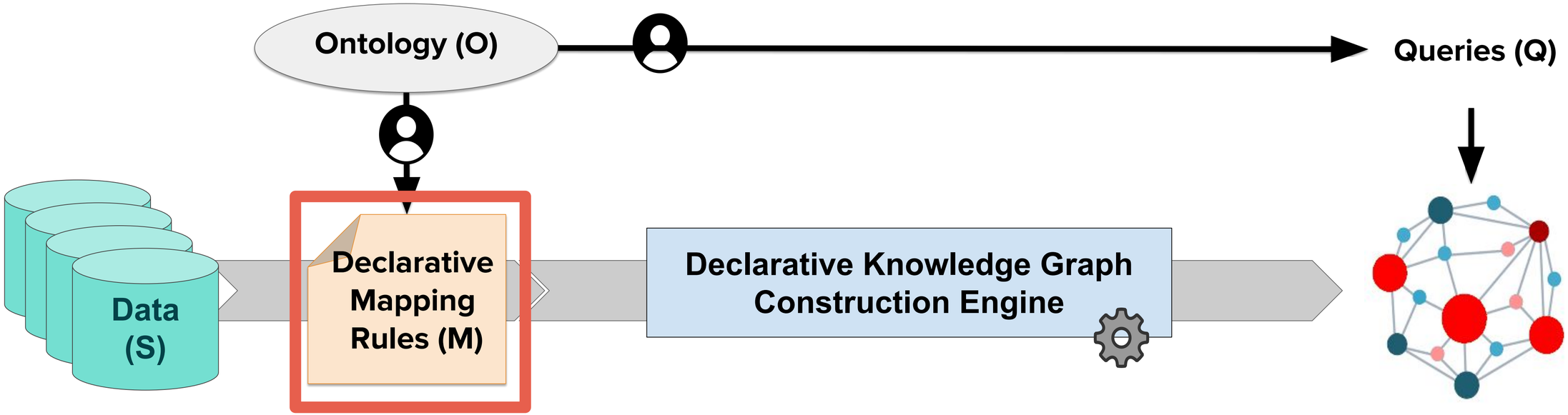
The construction of KG (v2.0)

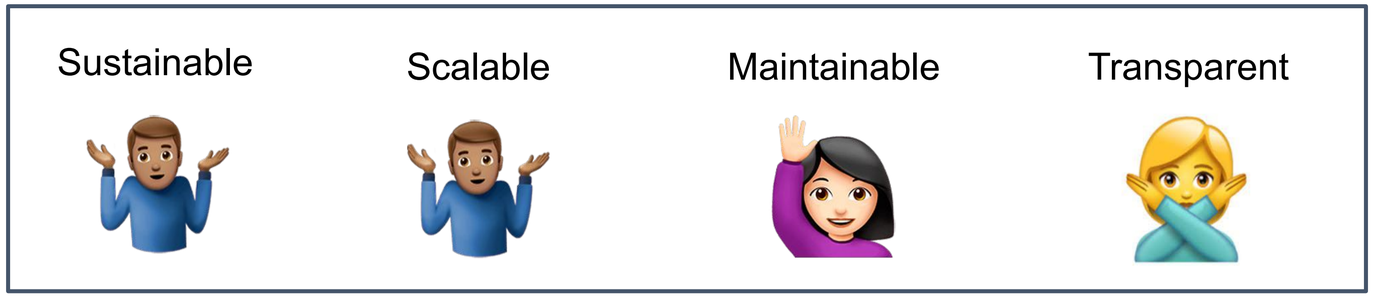


Knowledge Graph Construction = Data Integration System (DIS) = <S, M, O>

Poggi, A., Lembo, D., Calvanese, D., De Giacomo, G., Lenzerini, M., & Rosati, R. (2008). Linking data to ontologies. In *Journal on data semantics X*
Lenzerini, M. Data integration: A theoretical perspective. In *Proceedings of the 21st ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*

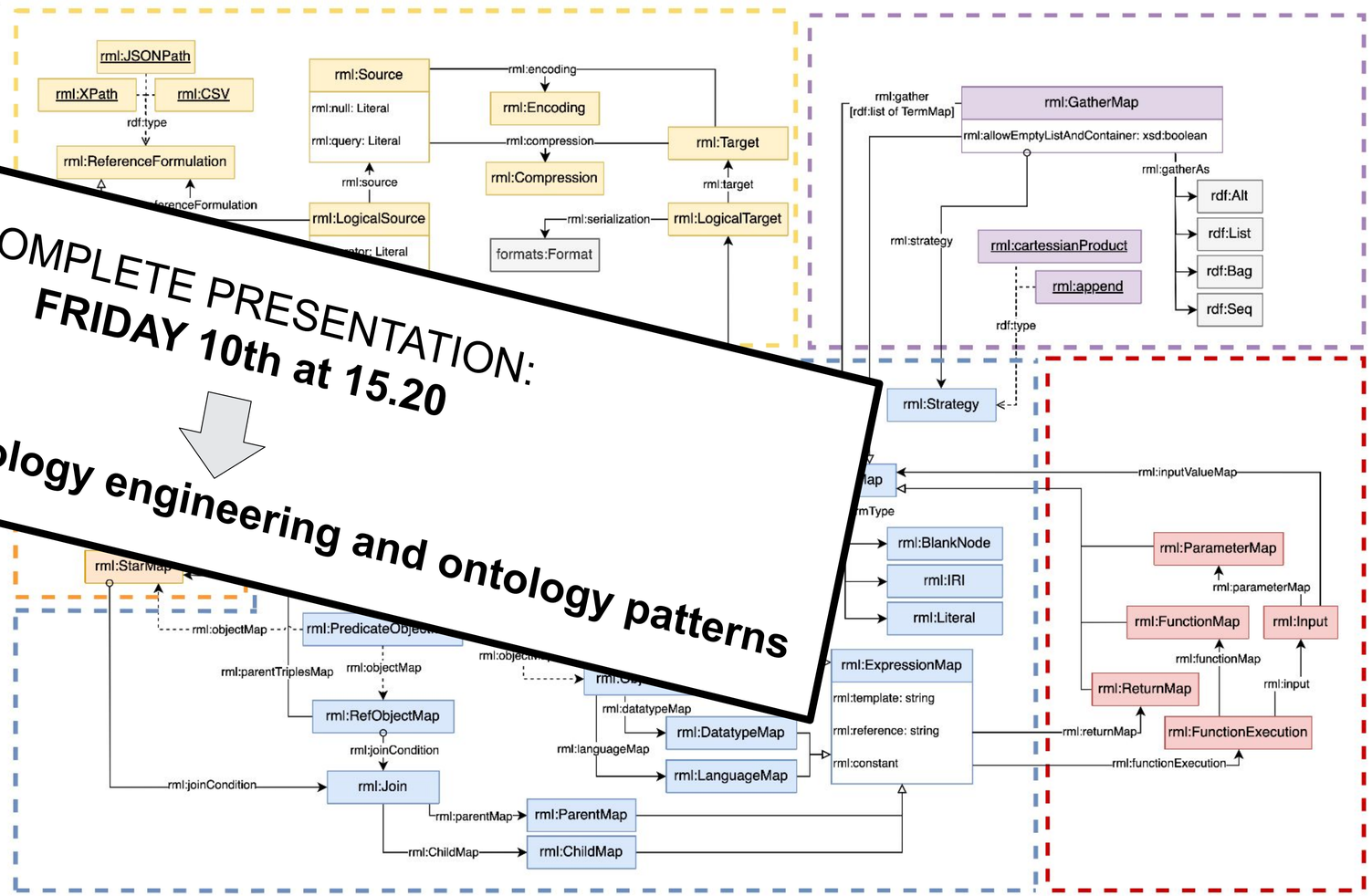
The construction of KG (v2.0)





The RML Ontology

- Five on-going spec
  - RML-Core: Sche
  - RML-IO: Source
  - RML-CC: Colle
  - RML-FNML: Da
  - RML-star: RDF
- Modular approach
- Unification of prefixes
- >15 people actively involved

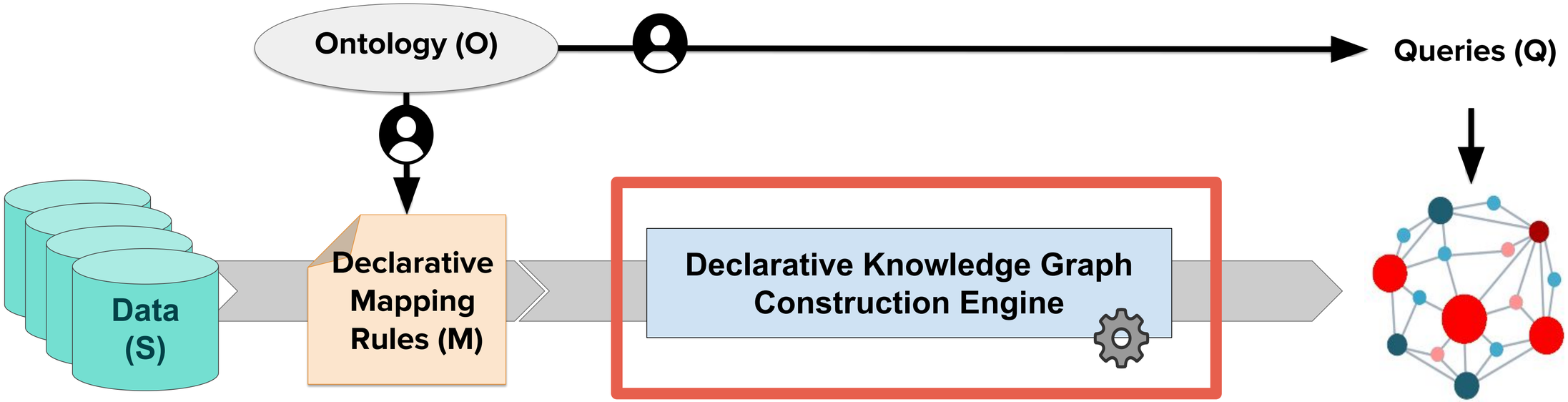


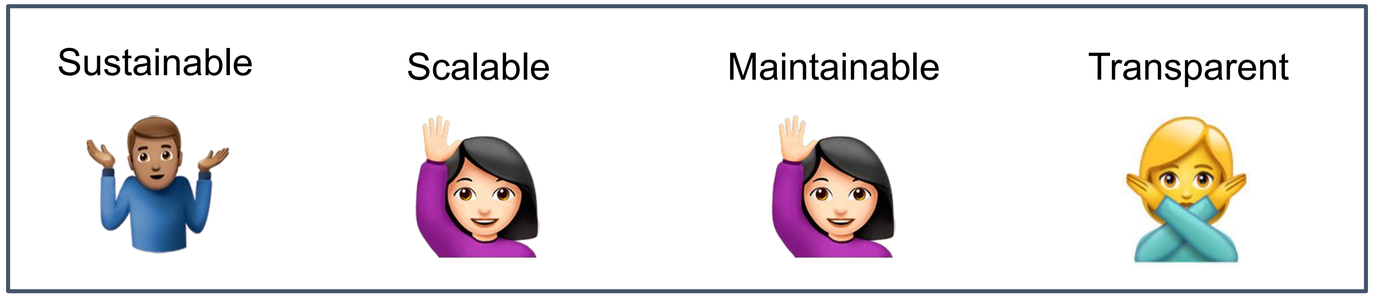
COMPLETE PRESENTATION:
FRIDAY 10th at 15.20

Session 10A : Ontology engineering and ontology patterns

Iglesias-Molina, A., Van Assche, D., Arenas-Guerrero, J., De Meester, B., Debruyne, C., Jozashoori, S., Maria, P., Michel, F., Chaves-Fraga, D., & Dimou, A. (2023). The RML Ontology: A Community-Driven Modular Redesign After a Decade of Experience in Mapping Heterogeneous Data to RDF. In *Proceedings of the 22nd International Conference on Semanitc Web.*

The construction of KG (v2.0)





Scalability in KG Construction: Tools & Benchmarks

Calvanese, D., Cogrel, B., Komla-Ebri, S., Kontchakov, R., Lanti, D., Rezk, M., ... & Xiao, G. (2017). Ontop: Answering SPARQL queries over relational databases. *Semantic Web, 8(3), 471-487.*

Chaves-Fraga, D., Ruckhaus, E., Priyatna, F., Vidal, M. E., & Corcho, O. (2021). Enhancing Virtual Ontology Based Access over Tabular Data with Morph-CSV. S*emantic Web Journal.*

Iglesias, E., Jozashoori, S., Chaves-Fraga, D., Collarana, D., & Vidal, M. E. (2020). SDM-RDFizer: An RML interpreter for the efficient creation of rdf knowledge graphs. In *Proceedings of the 29th ACM CIKM.*

Jozashoori, S., Chaves-Fraga, D., Iglesias, E., Vidal, M. E., & Corcho, O. (2020, November). FunMap: Efficient Execution of Functional Mappings for Knowledge Graph Creation. In International Semantic Web Conference

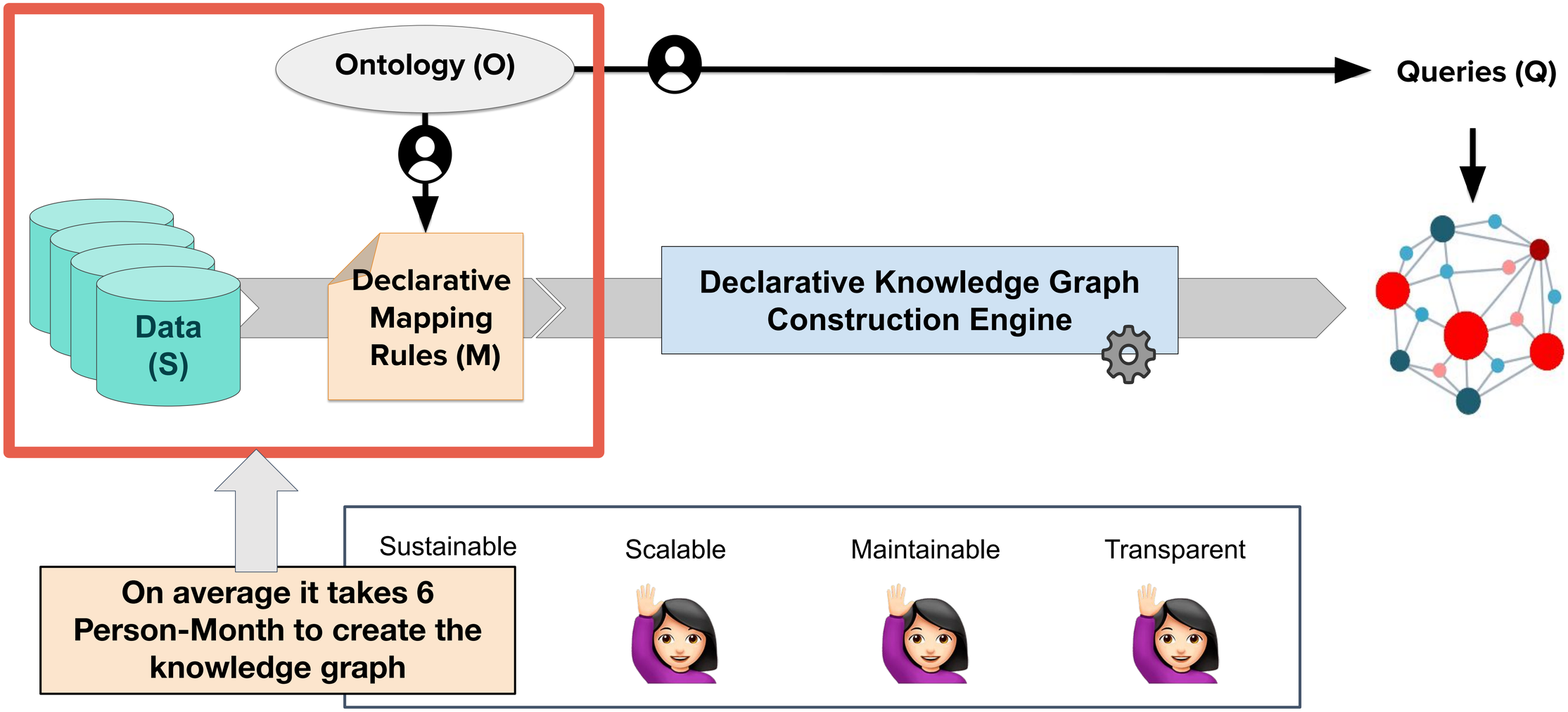
Arenas-Guerrero, J., Chaves-Fraga, D., Toledo, J., Pérez, M. S., & Corcho, O. (2022). Morph-KGC: Scalable knowledge graph materialization with mapping partitions. *Semantic Web Journal.*

Iglesias, E., Jozashoori, S., & Vidal, M. E. (2023). Scaling up knowledge graph creation to large and heterogeneous data sources. *Journal of Web Semantics, 75, 100755.*

Chaves-Fraga, D., Priyatna, F., Cimmino, A., Toledo, J., Ruckhaus, E., & Corcho, O. (2020). GTFS-Madrid-Bench: A benchmark for virtual knowledge graph access in the transport domain. *Journal of Web Semantics.*

Van Assche, D., Iglesias-Molina, A., Chaves-Fraga, D., Dimou, A., and Serles, U., (2023). First Edition of the Knowledge Graph Construction Challenge - Scalability and Performance. In *Knowledge Graph Construction Workshop - Extended Semantic Web Conference.*

The construction of KG (v2.0)



Human-Friendly KG Construction?

YARRRML

Unofficial Draft 26 January 2023

Editors:

Dylan Van Assche, Ghent University - IDLab, imec, dylan.vanassche@ugent.be
Ben De Meester, Ghent University - IDLab, imec, ben.demeester@ugent.be
Pieter Heyvaert, Ghent University - IDLab, imec, pieter.heyvaert@ugent.be
Anastasia Dimou, Ghent University - IDLab, imec, anastasia.dimou@ugent.be

This Version:

https://w3id.org/yarrrml/spec/%thisDate%/

Previous Version:

https://w3id.org/yarrrml/spec/%prevDate%/

Website:

https://w3id.org/yarrrml

Abstract

YARRRML (pronounced /jɑɹməl/) is a human readable text-based representation for declarative generation rules. It is a subset of [YAML], a widely used data serialization language designed to be human-friendly.

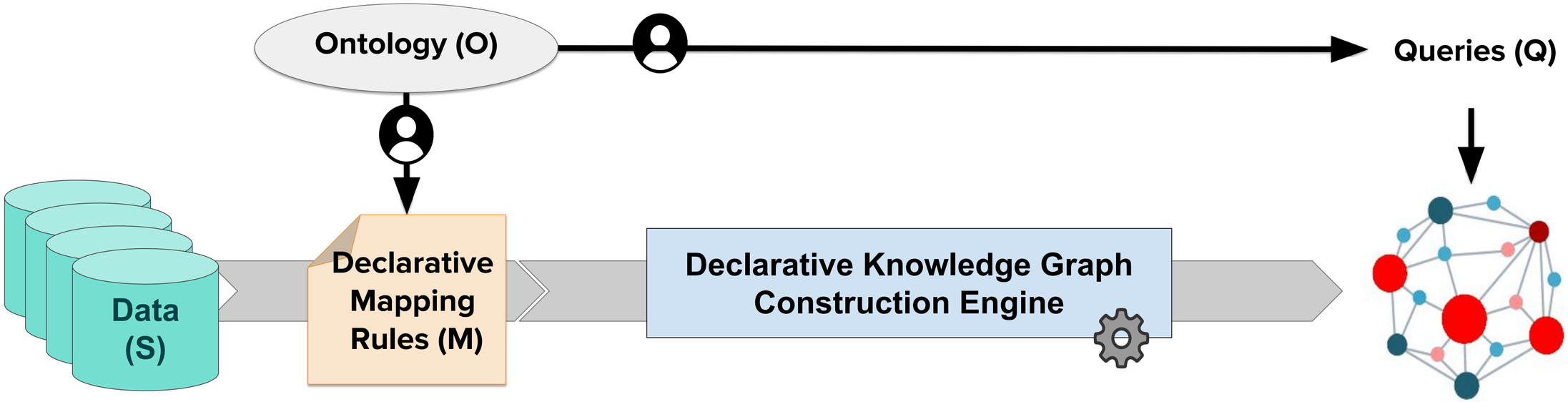
YATTER

MAPEATHOR

Iglesias-Molina, A., Chaves-Fraga, D., Dasoulas, I., & Dimou (2023), A. Human-Friendly Knowledge Graph Construction: Which one do you chose?. *International Conference on Web Engineering.*

Iglesias-Molina, A., Pozo-Gilo, L., Dona, D., Ruckhaus, E., Chaves-Fraga, D., & Corcho, O. (2020, November). Mapeathor: Simplifying the Specification of Declarative Rules for Knowledge Graph Construction. In *International Semantic Web Conference* (P&D).

The construction of KG (v2.0)



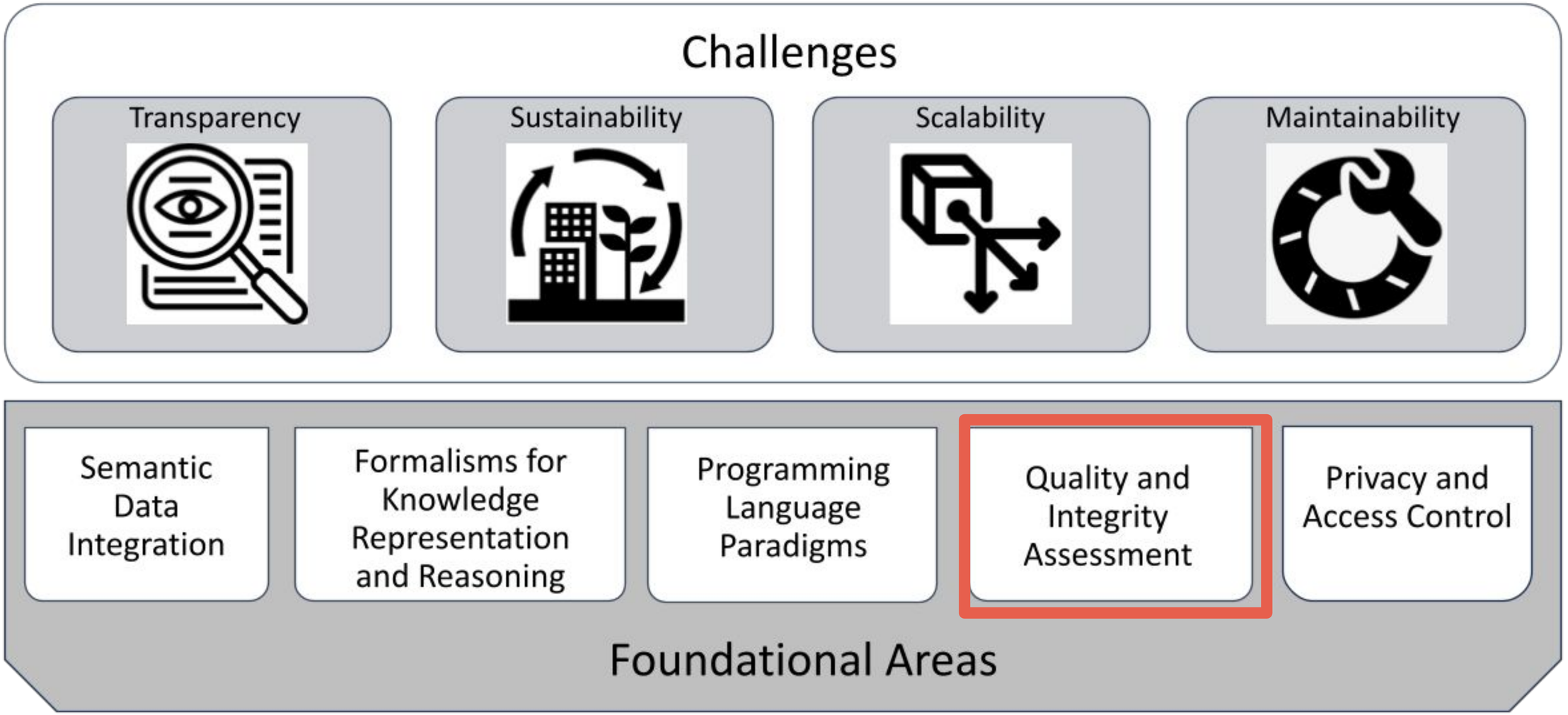
Sustainable

Scalable

Maintainable

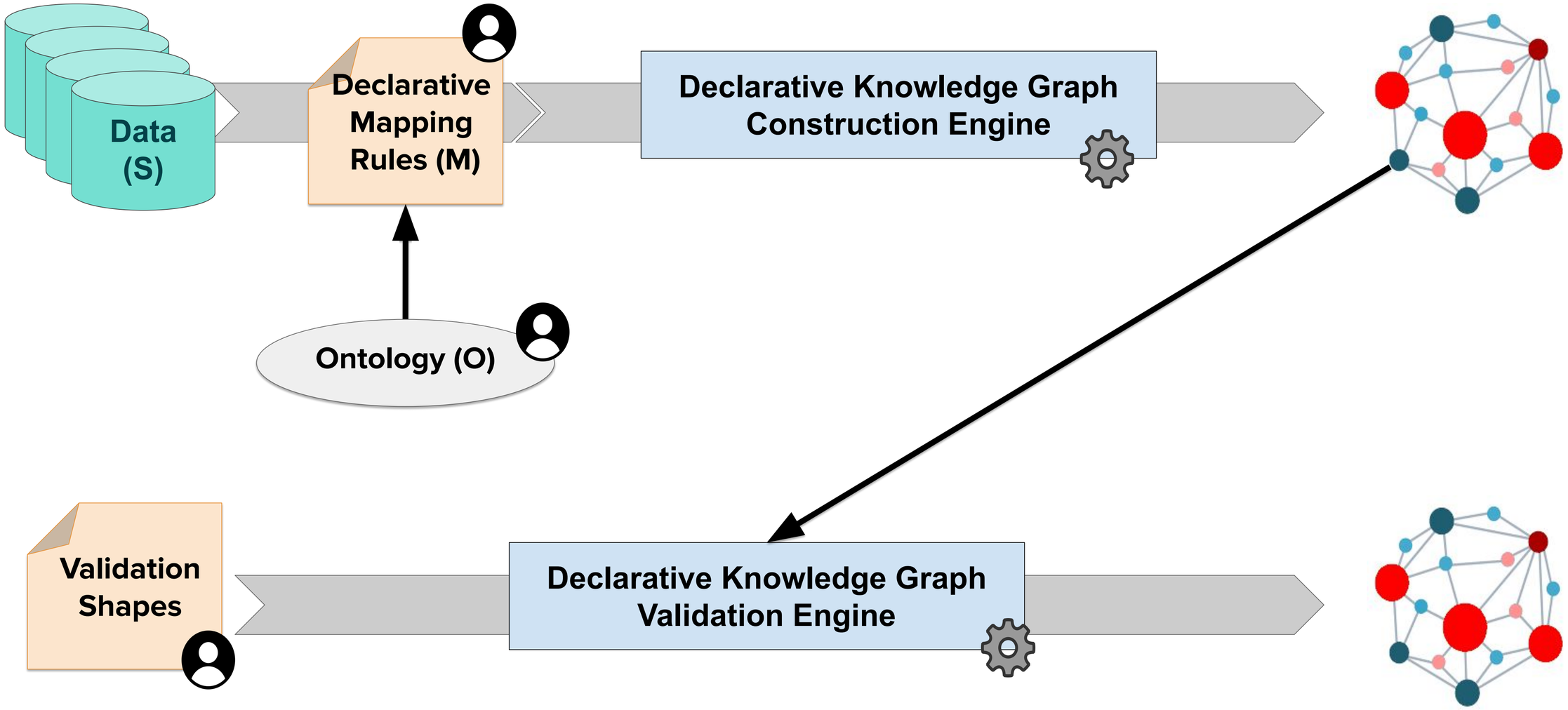
Transparent

Let’s make data management “sexy”…



Chaves-Fraga, D., Corcho, O., Dimou, A., & Vidal ME. (2023). Challenges on Data Management for Knowledge Graphs (Under Review)

A KG-driven data ecosystem (v0.1)



Validating KGs

Ghiasnezhad Omran, P., Taylor, K., Rodríguez Méndez, S., & Haller, A. (2023). Learning SHACL shapes from knowledge graphs. *Semantic Web, 14(1), 101-121.*

Cimmino, A., Fernández-Izquierdo, A., & García-Castro, R. (2020, May). Astrea: automatic generation of shacl shapes from ontologies. In *European Semantic Web Conference (pp. 497-513). Cham: Springer International Publishing.*

Fernandez-Álvarez, D., Labra-Gayo, J. E., & Gayo-Avello, D. (2022). Automatic extraction of shapes using sheXer. *Knowledge-Based Systems, 238, 107975.*

Delva, T., De Smedt, B., Min Oo, S., Van Assche, D., Lieber, S., & Dimou, A. (2021, December). Rml2shacl: Rdf generation taking shape. In *Proceedings of the 11th on Knowledge Capture Conference (pp. 153-160).*

Garcia-Gonzalez, H., & Labra-Gayo, J. E. (2020). XMLSchema2ShEx: Converting XML validation to RDF validation. *Semantic Web, 11(2), 235-253.*

Rabbani, K., Lissandrini, M., & Hose, K. (2023). Extraction of Validating Shapes from very large Knowledge Graphs. *Proceedings of the VLDB Endowment, 16(5), 1023-1032.*

Wright, J., Rodríguez Méndez, S. J., Haller, A., Taylor, K., & Omran, P. G. (2020, November). Schímatos: a SHACL-based web-form generator for knowledge graph editing. *In International Semantic Web Conference (pp. 65-80). Cham: Springer International Publishing.*

Figuera, M., Rohde, P. D., & Vidal, M. E. (2021, April). Trav-SHACL: Efficiently validating networks of SHACL constraints. In *Proceedings of the Web Conference 2021 (pp. 3337-3348).*

Trav-SHACL

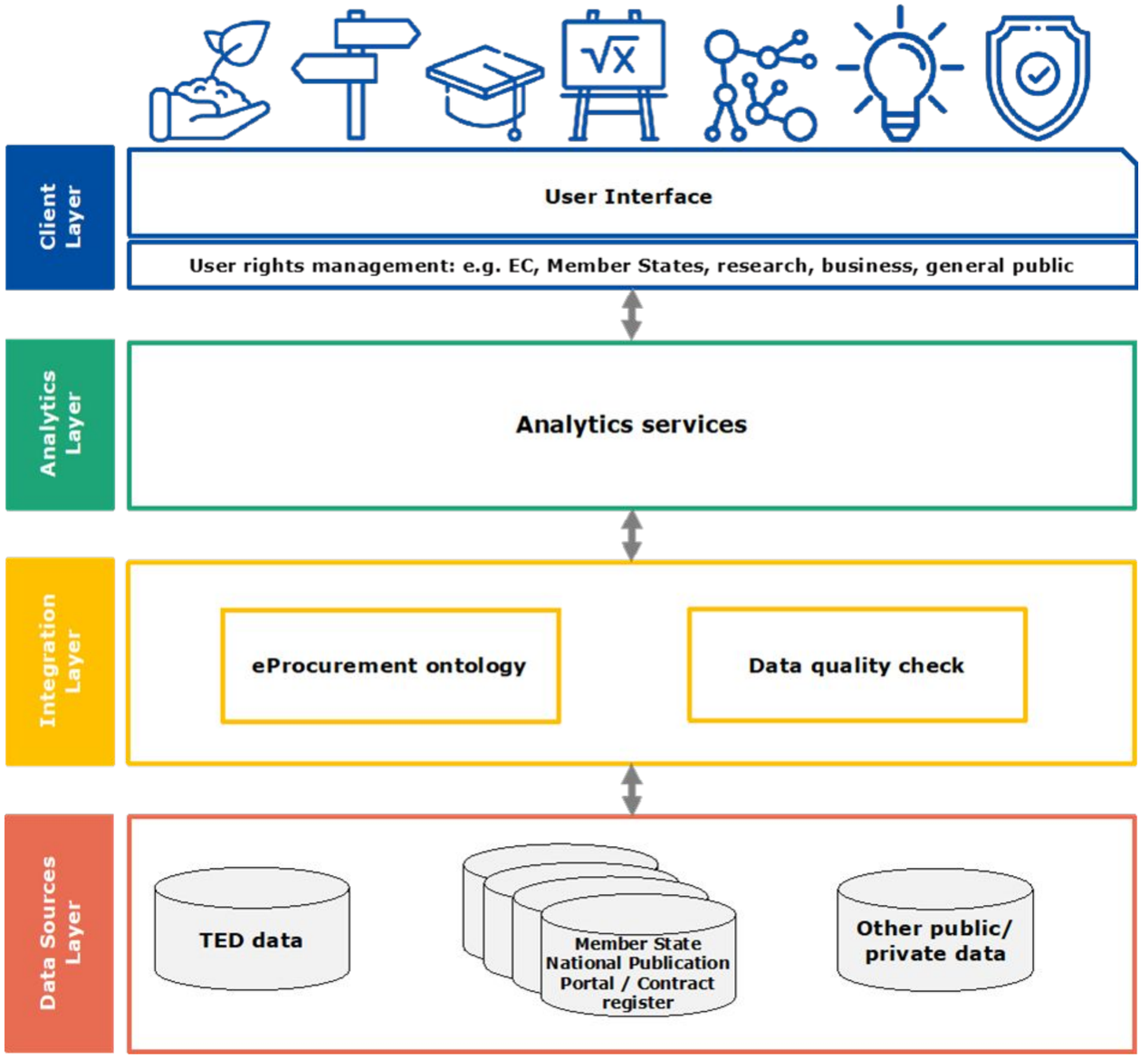
The EU Public Procurement Data Space

Homogenize the access to all public procurement data across Europe

Calculate standard transparency indicators for each member state

Researching on:

- vocabularies,
- **resources maintainability,**
- federated query processing,
- **semantic data ingestion,**
- query performance/scalability...



European Commission

Ontology Engineering Group

NTT DATA

CIRUS Singular Research Center on Intelligent technologies

The scenario: Nothing under your control

The e-Procurement Ontology (ePO)

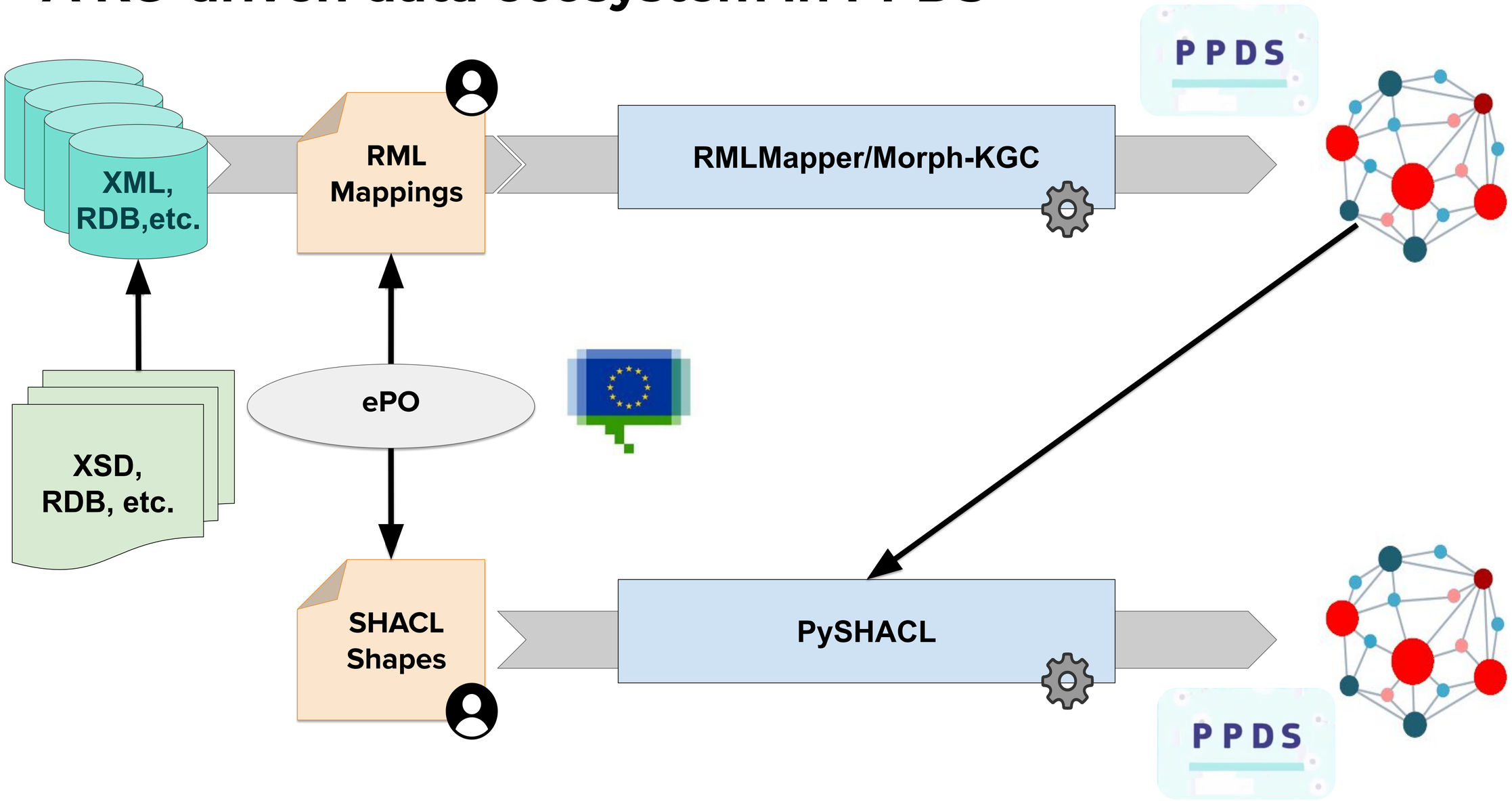
- Developed and maintained by the EU Publication Office
- Not 100% stable (releases every ~6 months)
- Partial support for transformation from TED (XML) to RDF using RML
- Complex workflows for generating the ontology and shapes (from UML…)

The EU Public Procurement Data Space:

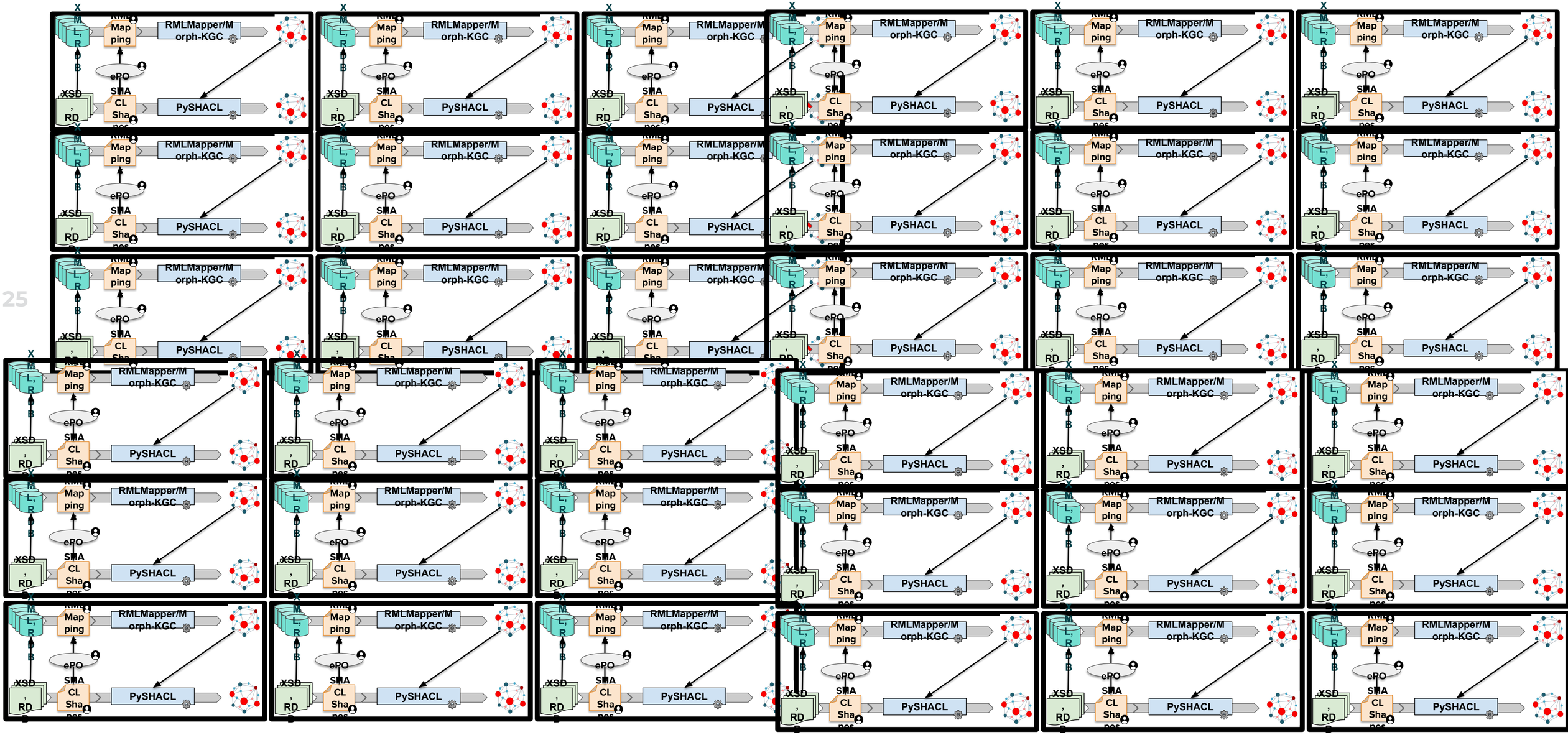
- Reuse the e-Procurement Ontology
- Support to all EU member states to make data compliant with ePO (RML)
- Ensure long term maintenance for all involved assets
- Ensuring efficient construction of knowledge graphs

Guasch, C., Lodi, G., & Dooren, S. V. (2022, October). Semantic knowledge graphs for distributed data spaces: The public procurement pilot experience. In *International Semantic Web Conference (pp. 753-769). Cham: Springer International Publishing*.

A KG-driven data ecosystem in PPDS



A KG-driven data ecosystem in PPDS (27 workflows)



A bit of Governance... the PPDS Conventions

PPDS Conventions

This document contains all PPDS data governance conventions. Their current master reference is here: PPDS data governance conventions ↗ In case of discrepancy, the master text should be treated as the statement of the convention. There is also the place to provide comments.

Ontology

- ONTO1. URIs for ontology terms
- ONTO2. Naming conventions for ontology terms
- ONTO3. Ontology term URI persistency
- ONTO4. Documentation for ontology terms
- ONTO5. Ontology publication following FAIR principles
- ONTO6. Ontology versioning
- ONTO7. Ontology design principles

Mapping

- MAP1. Transformation rules from original data sources into RDF need to be expressed declaratively

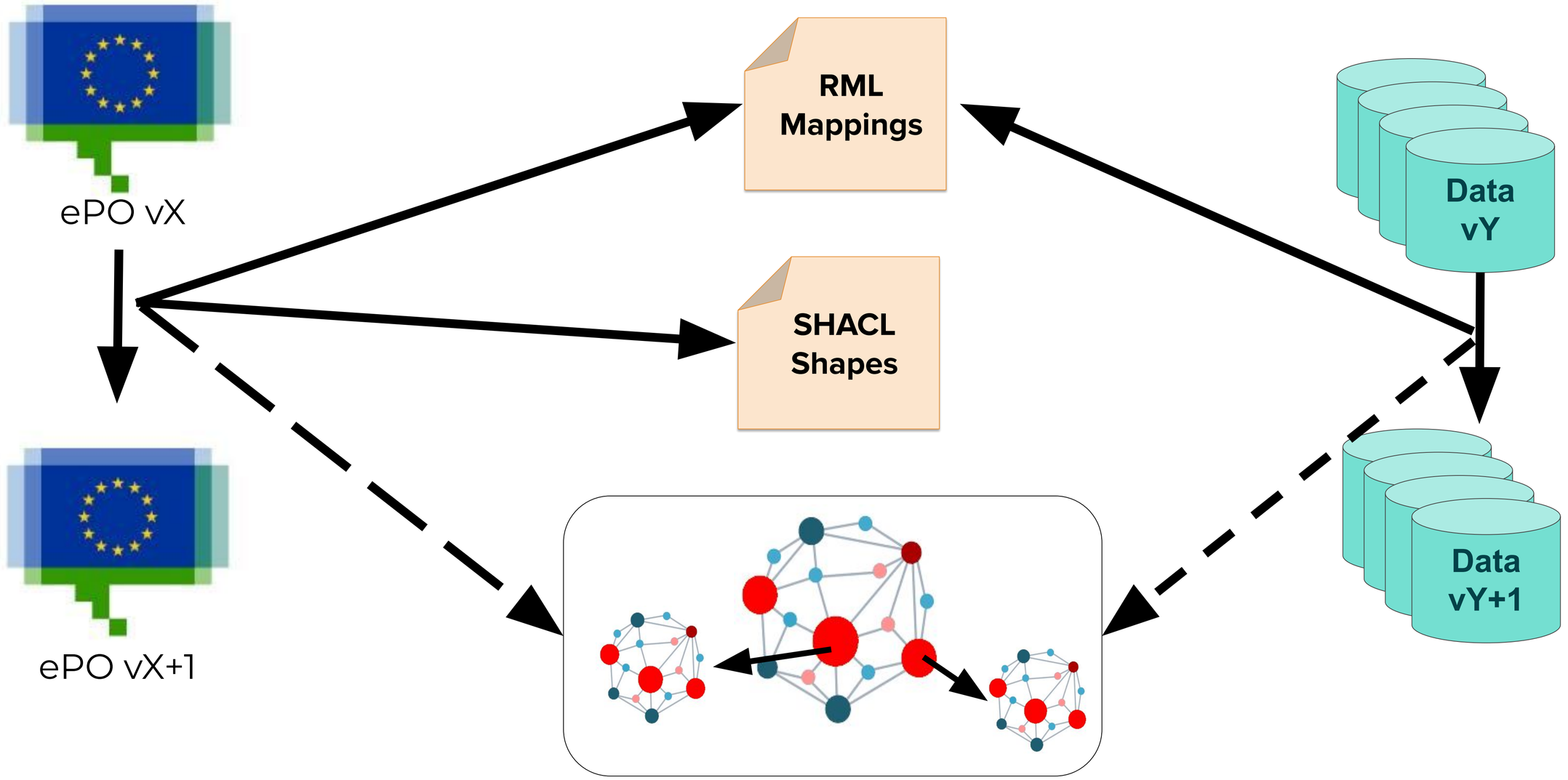
Shapes

- SHA1. Data validation must be done using SHACL shapes
- SHA2. URIs for SHACL Shapes

Graph

- GRAPH1. URIs for named graphs
- GRAPH2. Documentation on named graphs

How to manage changes? Ontology, data, metadata



Open questions...

Can we minimize the impact (w.r.t. a decentralized KG) of

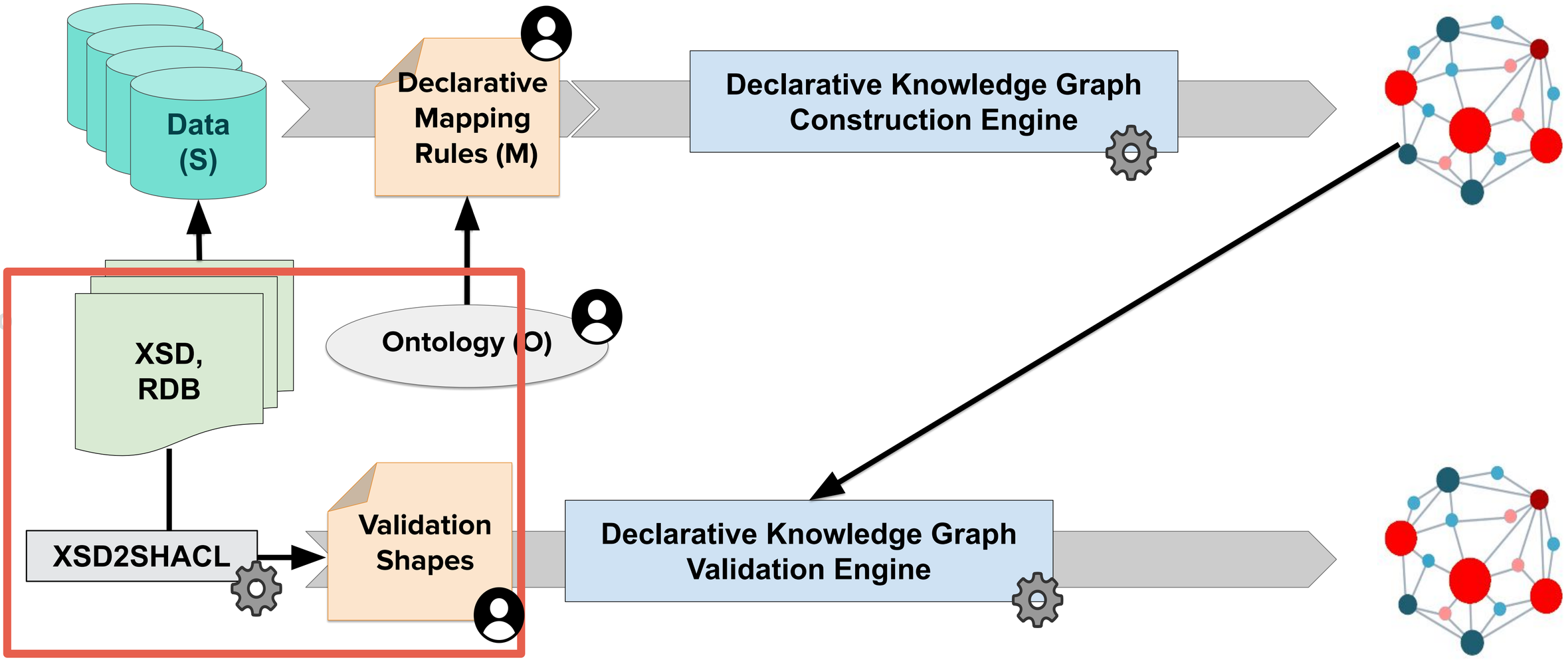
- data source changes?
- metadata representation model?
- the ontology changes?

Open questions...

Can we minimize the impact (w.r.t. the decentralized KG) of

- **data source changes?**
- metadata representation model?
- the ontology changes?

Data Constraints Evolution



Duan, X., Chaves-Fraga, D., & Dimou, A. (2023). XSD2SHACL: Capturing RDF Constraints from XML Schema. *In International Conference on Knowledge Capture (K-CAP)*. Code available at https://github.com/dtai-kg/XSD2SHACL

XSD2SHACL: Approach

XML Schema	SHACL
Complex type *element-only content & empty content* <xs:element name="*N*"> <xs:element name="*N*" type="*C*"/>	a sh:NodeShape sh:targetClass :*N*
Complex type *simple content & mixed content* <xs:element name="*N*"> <xs:element name="*N*" type="*C*"/>	a sh:NodeShape sh:targetClass :*N* a sh:PropertyShape sh:path :*N*
Simple type: user-defined <xs:element name="*N*"> <xs:attribute name="*N*"> <xs:element name="*N*" type="*C*"/> <xs:attribute name="*N*" type="*C*"/>	a sh:PropertyShape sh:path :*N*
Simple type: built-in <xs:element name="*N*" type="*C*"/> <xs:attribute name="*N*" type="*C*"/>	a sh:PropertyShape sh:path :*N* sh:datatype *C*
<xs:complexType name="*N*"> <xs:group name="*N*"> <xs:attributeGroup name="*N*">	a sh:NodeShape
xs:sequence	sh:order
xs:choice	sh:xone
xs:union	sh:or
xs:annotation	sh:description

XML Schema	SHACL
Complex type C <xs:restriction base="*C*"/> <xs:extension base="*C*">	sh:node
Built-in simple type C <xs:extension base="*C*"/> <xs:restriction base="*C*"/>	sh:datatype *C*
<xs:pattern value="*C*"/>	sh:pattern *C*
<xs:minLength value="*C*"/>	sh:minLength *C*
<xs:maxLength value="*C*"/>	sh:maxLength *C*
<xs:Length value="*C*"/>	sh:minLength *C* sh:maxLength *C*
<xs:minInclusive value="*C*"/>	sh:minInclusive *C*
<xs:maxInclusive value="*C*"/>	sh:maxInclusive *C*
<xs:minExclusive value="*C*"/>	sh:minExclusive *C*
<xs:maxExclusive value="*C*"/>	sh:maxInclusive *C*
<xs:enumeration value="*C1*"/> <xs:enumeration value="*C2*"/>	sh:in (*C1*, *C2*)

Duan, X., Chaves-Fraga, D., & Dimou, A. (2023). XSD2SHACL: Capturing RDF Constraints from XML Schema. *In International Conference on Knowledge Capture (K-CAP)*. Code available at https://github.com/dtai-kg/XSD2SHACL

XSD2SHACL: Results

Use Case	Target Declaration			Property Path		
	C_T	R/T	R/T'	C_P	R/P	R/P'
RINF						
contact-line	1	1/1	1/1	8	8/8	**10/10**
etcs-levels	1	1/1	1/1	2	2/2	**3/3**
meso-net-e	1	1/1	1/1	2	2/2	2/2
meso-net-r	1	1/1	1/1	2	2/2	2/2
op-tracks	1	1/1	1/1	11	11/11	**13/13**
operational	2	2/2	**5/5**	13	13/13	**23/23**
platforms	1	1/1	1/1	7	7/7	**9/9**
sections-of	1	1/1	1/1	9	9/9	**14/14**
sidings	1	1/1	1/1	16	16/16	**18/18**
sol-tracks	1	**1/2**	**1/1**	95	95/107	**110/110**
train-detect	1	1/1	1/1	24	24/26	**30/30**
tunnels	1	1/1	**3/3**	13	13/14	**21/21**
TED						
F03	50	50/278	50/**50**	129	132/493	**144/144**
F06	50	50/278	50/**50**	129	134/493	**144/144**
F13	50	50/278	50/**50**	127	132/493	**142/142**

Post-adjustment with RML for comparing results w.r.t. human-defined SHACL

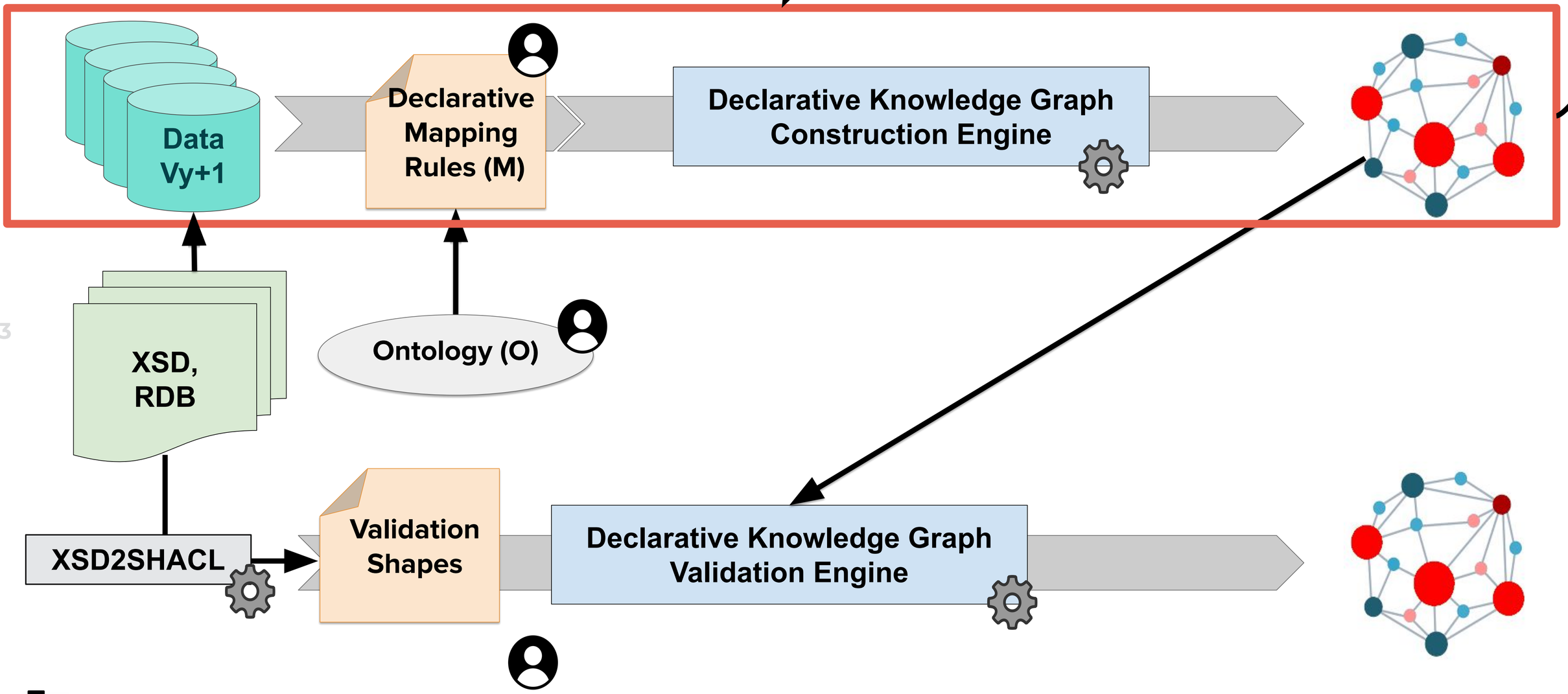
Two use cases: PPDS (TED data) and RINF (EU Railway Agency - ERA)

Most of the shapes are correctly extracted (legacy XSD or SHACL shapes that do not validate anything)

Performance:
- PPDS: 30 XSD files → 1144 Shapes ~20 secs
- ERA: 1 XSD file → 198 Shapes <1 sec

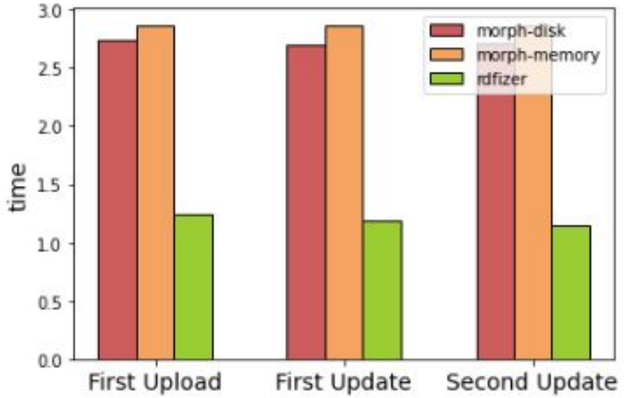
Duan, X., Chaves-Fraga, D., & Dimou, A. (2023). XSD2SHACL: Capturing RDF Constraints from XML Schema. *In International Conference on Knowledge Capture (K-CAP)*. Code available at https://github.com/dtai-kg/XSD2SHACL

Re-Constructing the KGs

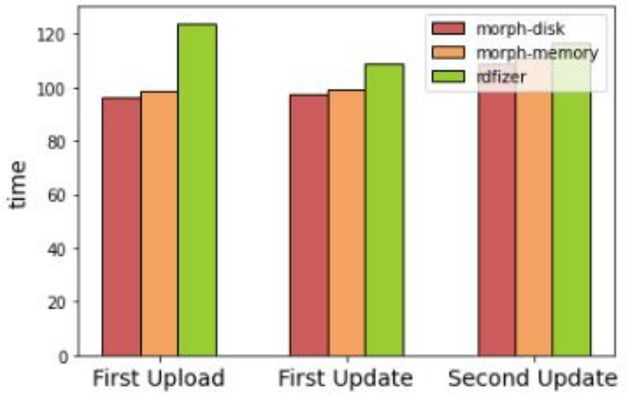


Cerezo, J, Incremental Knowledge Graph Construction. Master in Data Science - UPM. Supervised by David Chaves-Fraga.

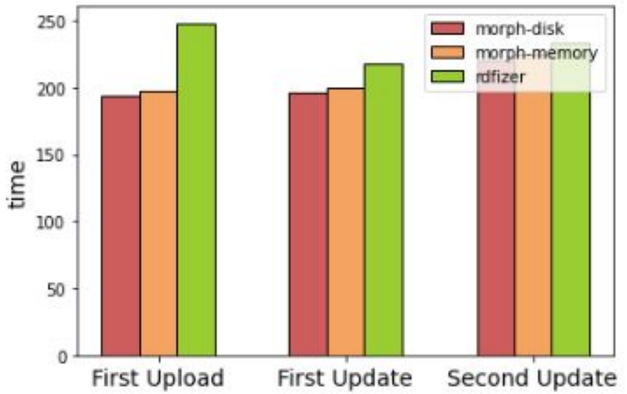
KG Construction without any optimization



GTFS-Bench Size 1

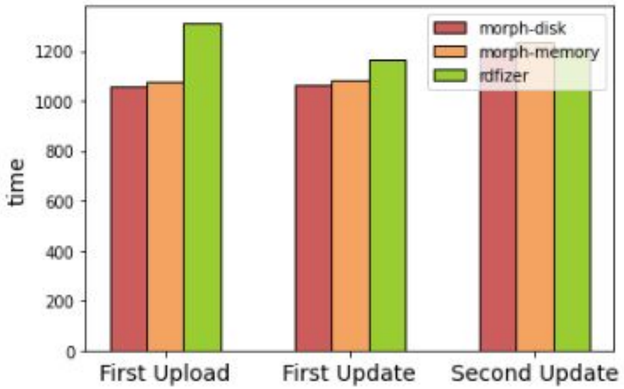


GTFS-Bench Size 5



GTFS-Bench Size 10

GTFS-Bench Size 25



GTFS-Bench Size 50

morph

SDM-RDFizer

GTFS-Madrid-Bench

*** The Spanish PPDS graph takes >12 hours to be constructed (without parallelism)**

First time in the KG construction

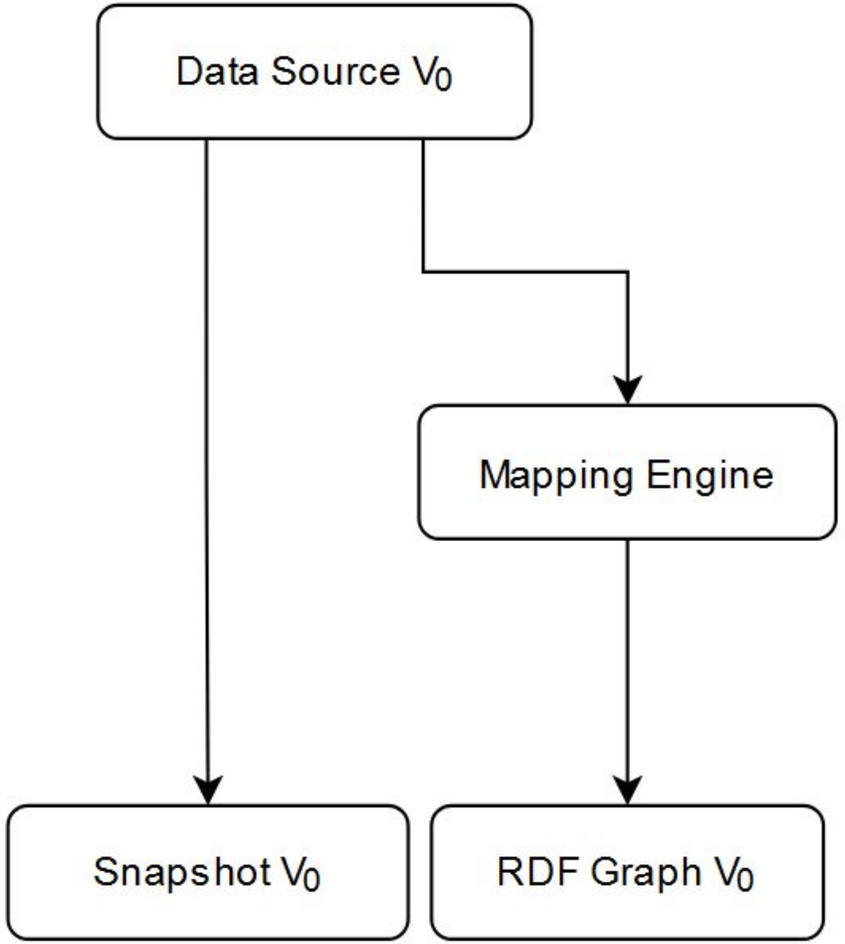
The data source is directly used to materialize the KG.

Input: the data source ($\mathbf{V_0}$).

Outputs: the KG ($\mathbf{V_0}$) and the snapshot ($\mathbf{V_0}$).

The snapshot is a file that represents a concrete version of the data source.

The execution time is the same as generating the KG from scratch.



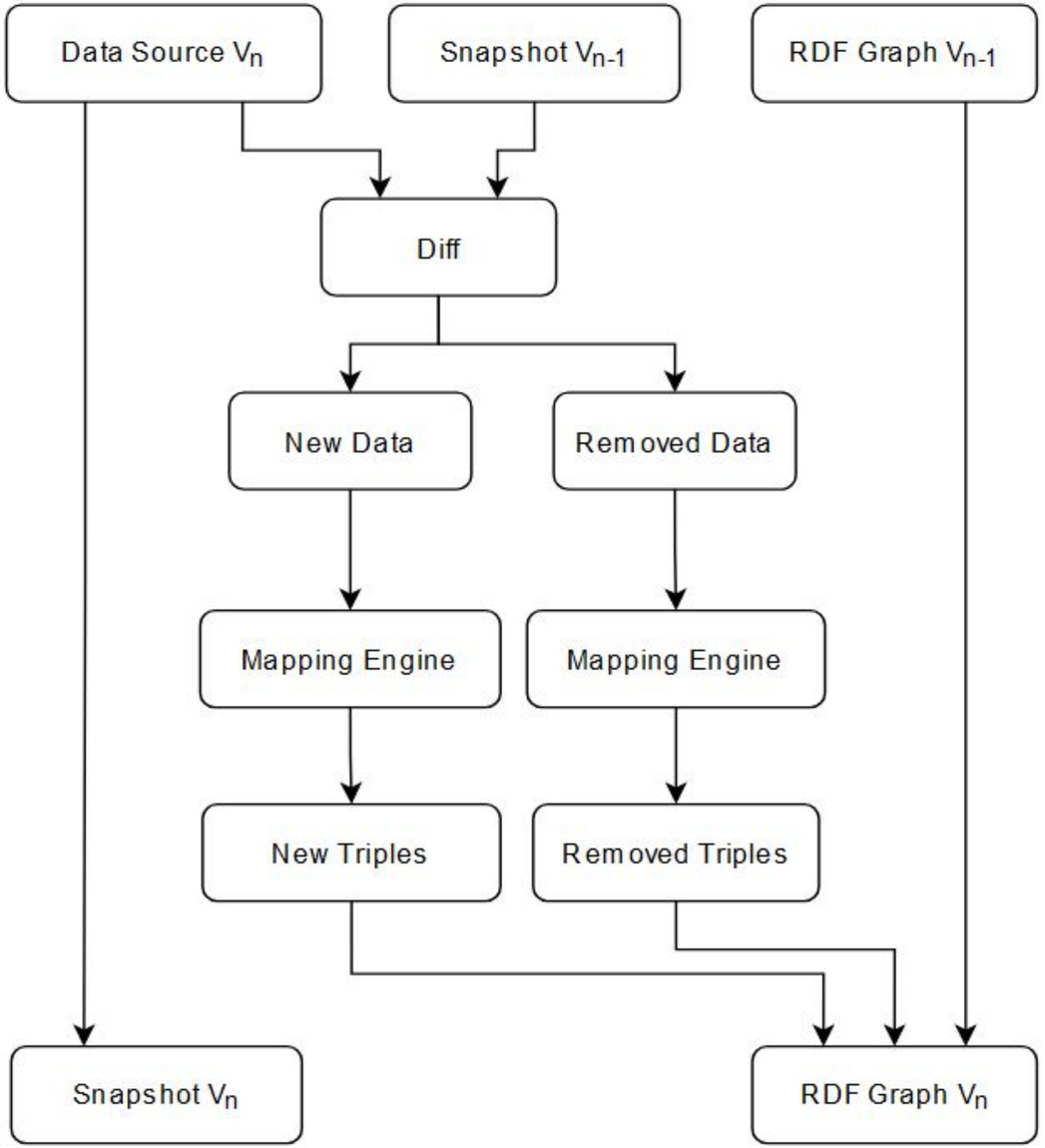
Updates during KG construction

Inputs:

- the data source (Vn)
- snapshot (Vn-1)
- KG (Vn-1).

Outputs:

- KG (Vn)
- Snapshot (Vn).



Setup

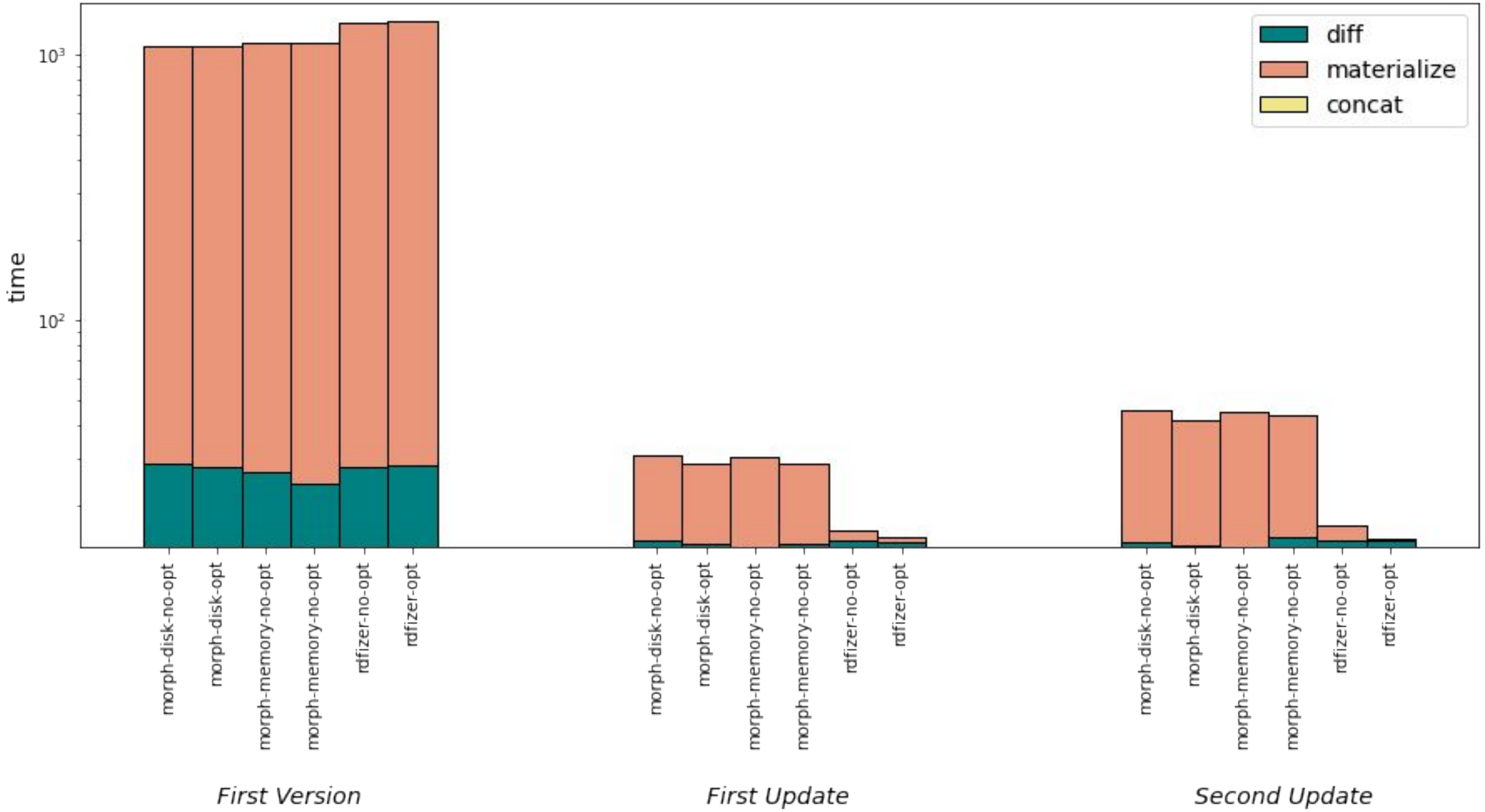
Approach: standard mapping engine vs. this proposal.
Benchmark: GTFS-Madrid-Bench size: 1, 5, 10, 25, 50.
Engines: Morph-KGC, SDM-RDFizer.
Mapping optimization: whether to select all or only the necessary mapping rules.
Engine configuration: whether to use in-memory data* source extension or not.
Additions/Deletions: 10% of the data.

Metrics:

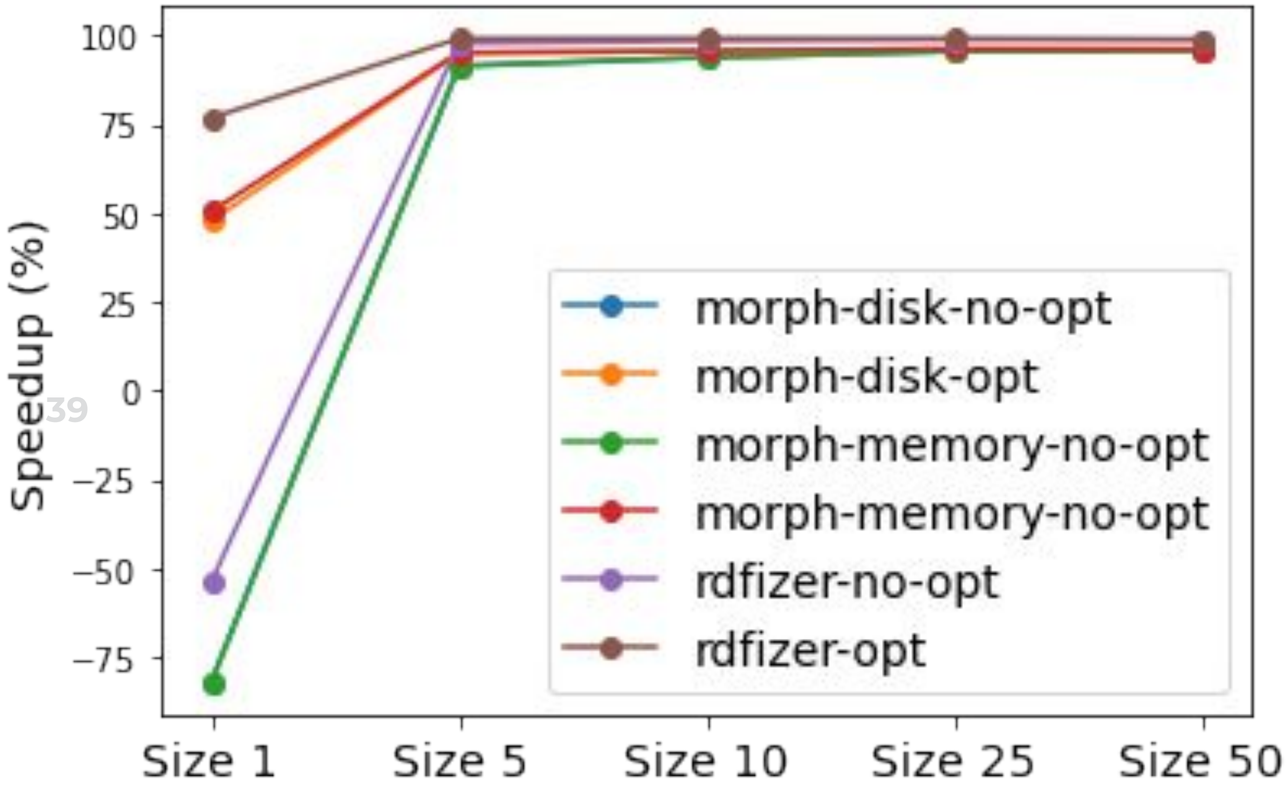
- ***Materialization* time**. The time elapsed during the materialization process.
- ***Diff* algorithm time**. The time elapsed during the execution of the diff algorithm.
- ***Merging time*.** The time elapsed during the construction of the final KG.

* Dasoulas, I., Chaves-Fraga, D., Garijo, D., & Dimou, A. Declarative RDF construction from in-memory data structures with RML. In: *Proceedings of the 4th International Workshop on Knowledge Graph Construction,* 2023.

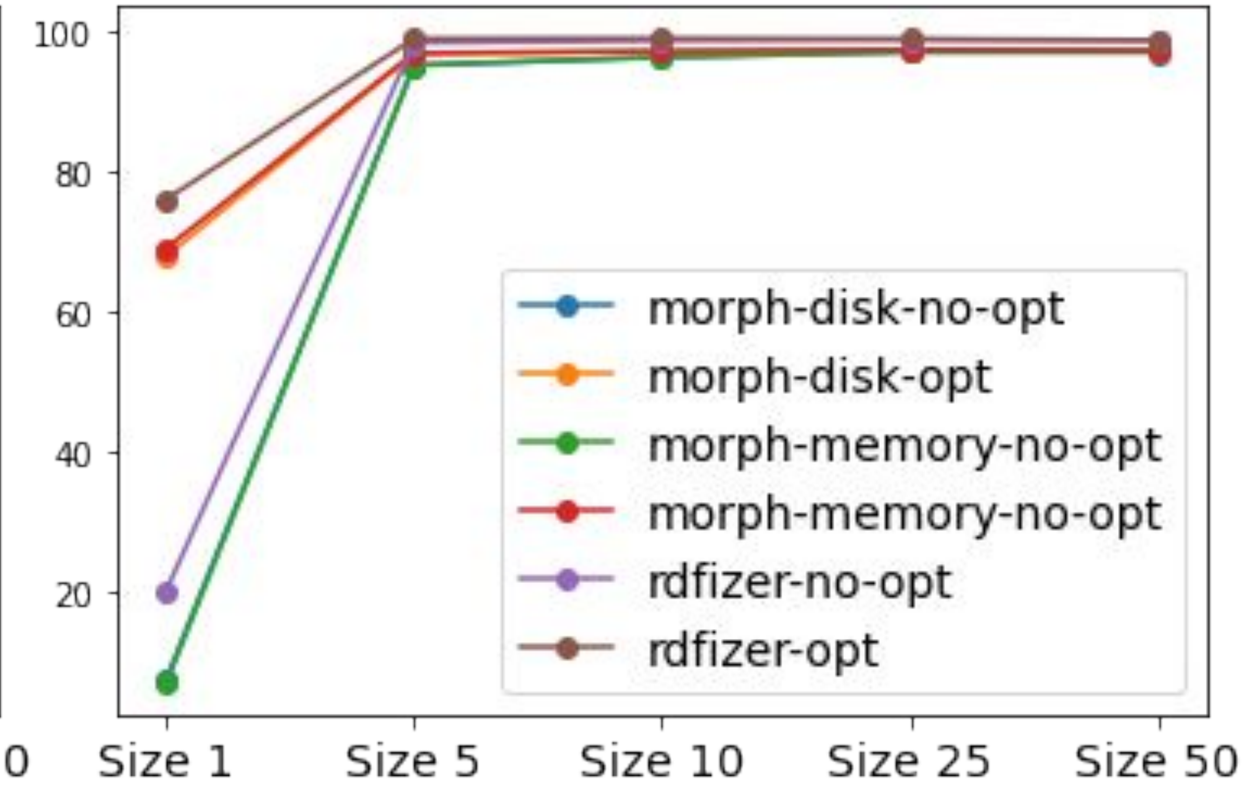
Handling updates in KG Construction (GTFS-50)



Handling updates in KG Construction



Adding new triples



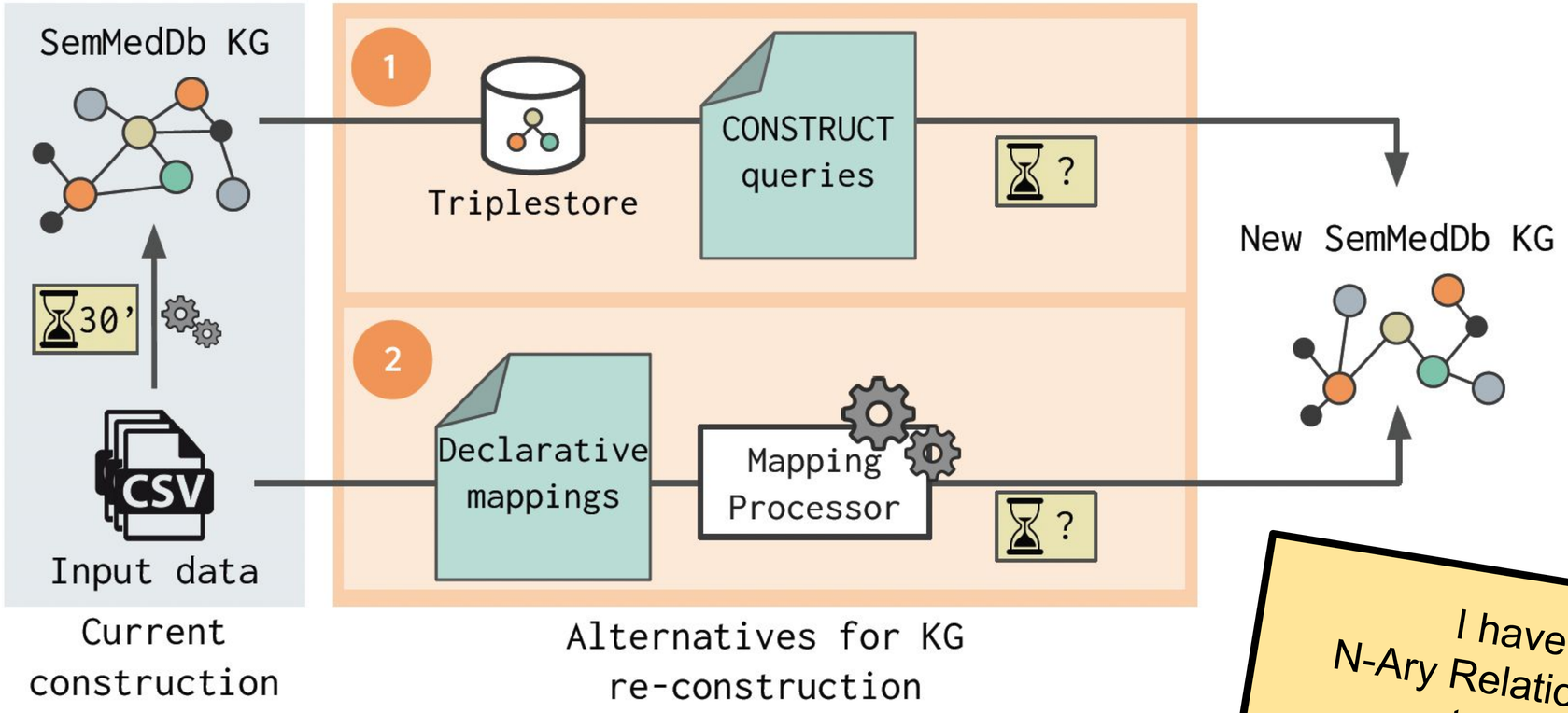
Updating triples

Open questions...

Can we minimize the impact (w.r.t. the decentralized KG) of

- data source changes?
- **metadata representation model?**
- the ontology changes?

Re-constructing a KG



I have a KG modelled with N-Ary Relationships and I want to have transform it into RDF-star

Iglesias Molina, A., Toledo, J., Corcho, O., & Chaves Fraga, D. (2023). Re-Construction Impact on Metadata Representation Models. *In International Conference on Knowledge Capture (K-CAP).*

Setup

Approach: triplestores VS KG construction engines

Benchmark: Semantic MEDLINE Database (SemMedDB) - 1K, 10K, 100K, 1M

Metadata Representations: Std. Reification, Named Graphs, N-Ary Rel, RDF-Star

Engines: Morph-KGC, SPARQL-Anything (KGC) / GraphDB, Fuseki, Oxigraph (SPARQL)

Mappings: 5 RML Mappings, 5 SPARQL-Anything Queries

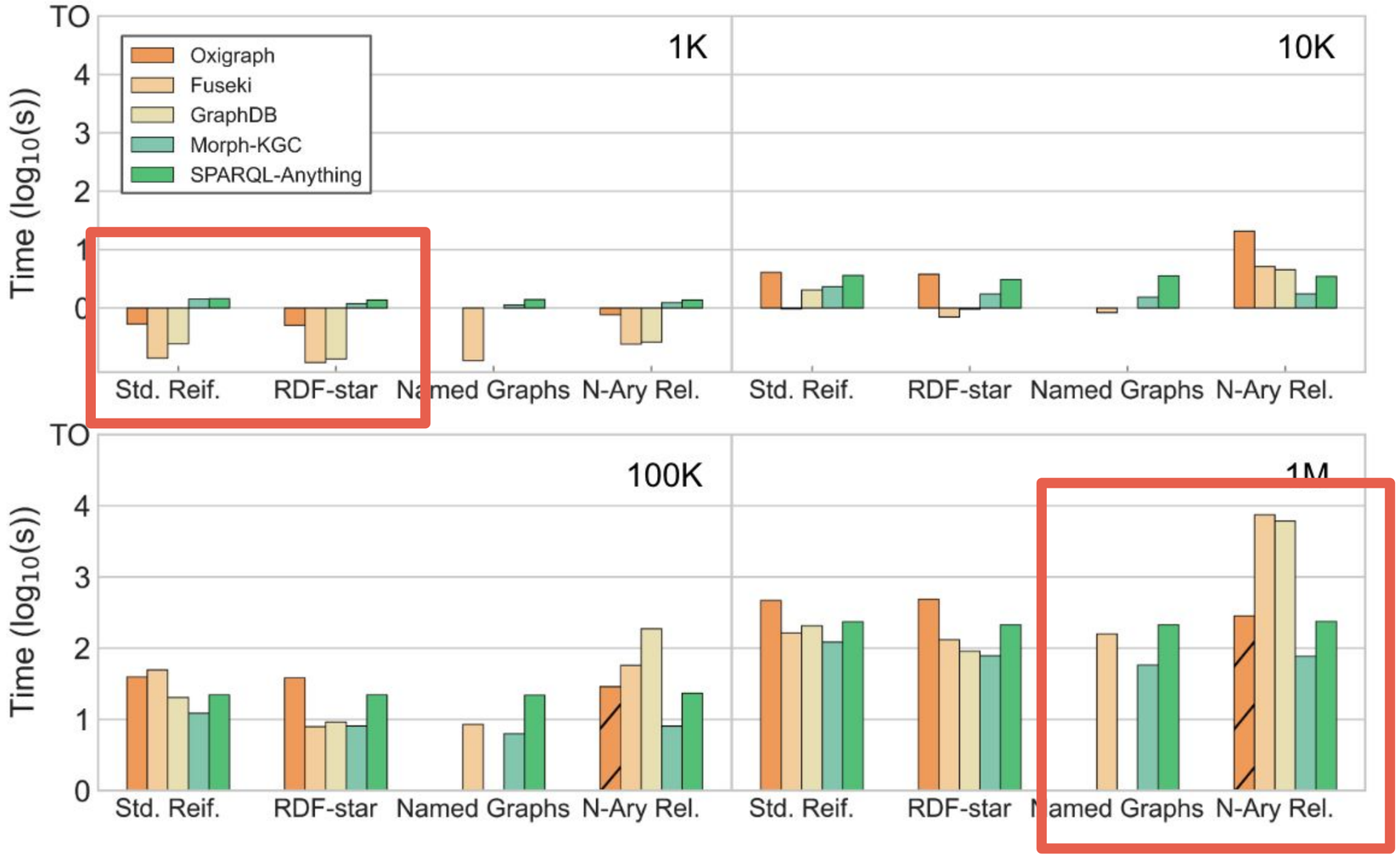
Queries: 12 Construct queries

Metrics:

- ***Materialization* time**: The time elapsed during the materialization process.
- ***Query execution* time**: The time elapsed during the executiong of the CONSTRUCT queries
- ***Geometric mean***: The central tendency of all execution times for a set of queries

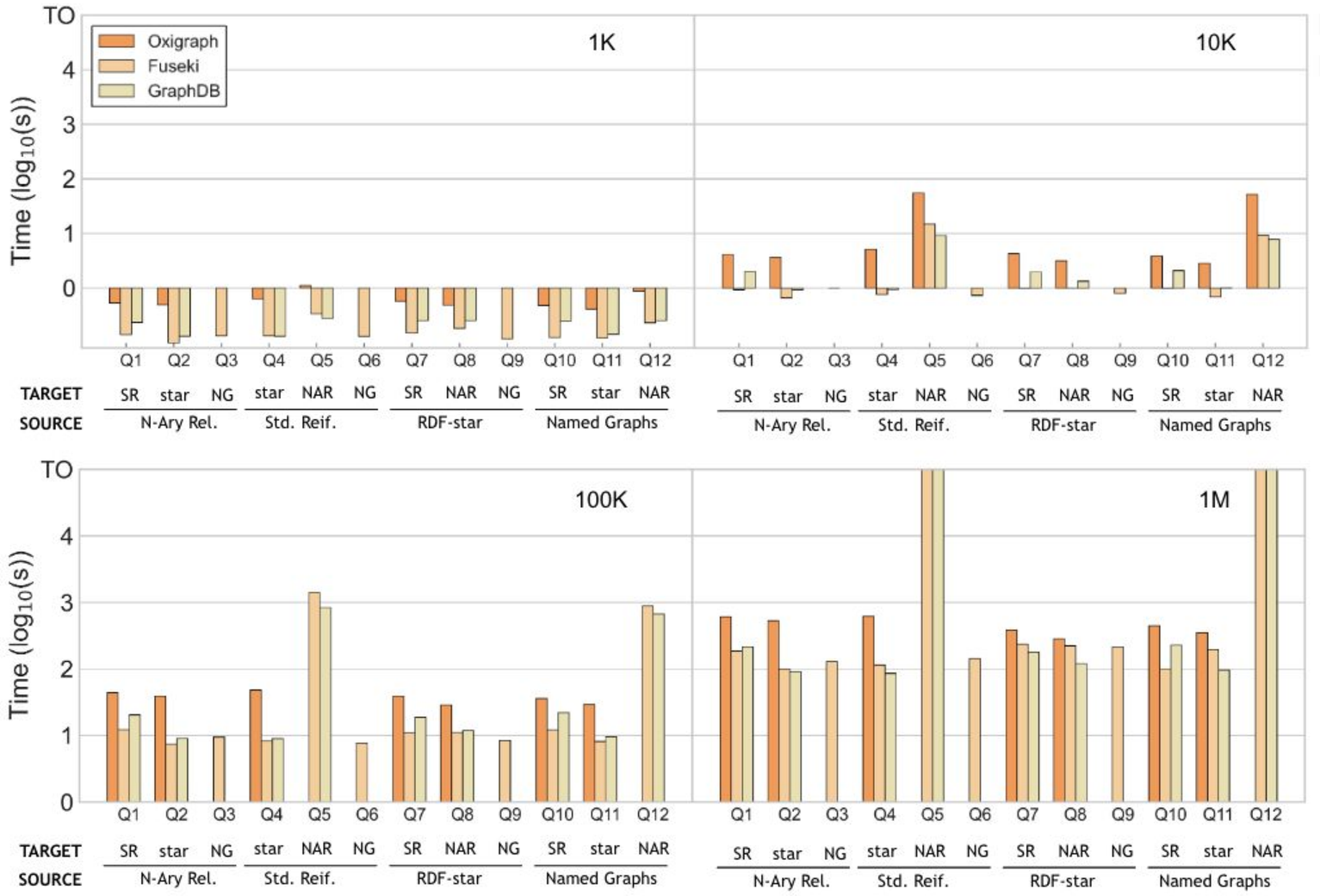
Iglesias Molina, A., Toledo, J., Corcho, O., & Chaves Fraga, D. (2023). Re-Construction Impact on Metadata Representation Models. *In International Conference on Knowledge Capture (K-CAP).*

KGC Engine VS triplestore for metadata change



Iglesias Molina, A., Toledo, J., Corcho, O., & Chaves Fraga, D. (2023). Re-Construction Impact on Metadata Representation Models. *In International Conference on Knowledge Capture (K-CAP).*

Triplestore performance for metadata change



Iglesias Molina, A., Toledo, J., Corcho, O., & Chaves Fraga, D. (2023). Re-Construction Impact on Metadata Representation Models. *In International Conference on Knowledge Capture (K-CAP).*

Open questions...

Can we minimize the impact (w.r.t. the decentralized KG) of

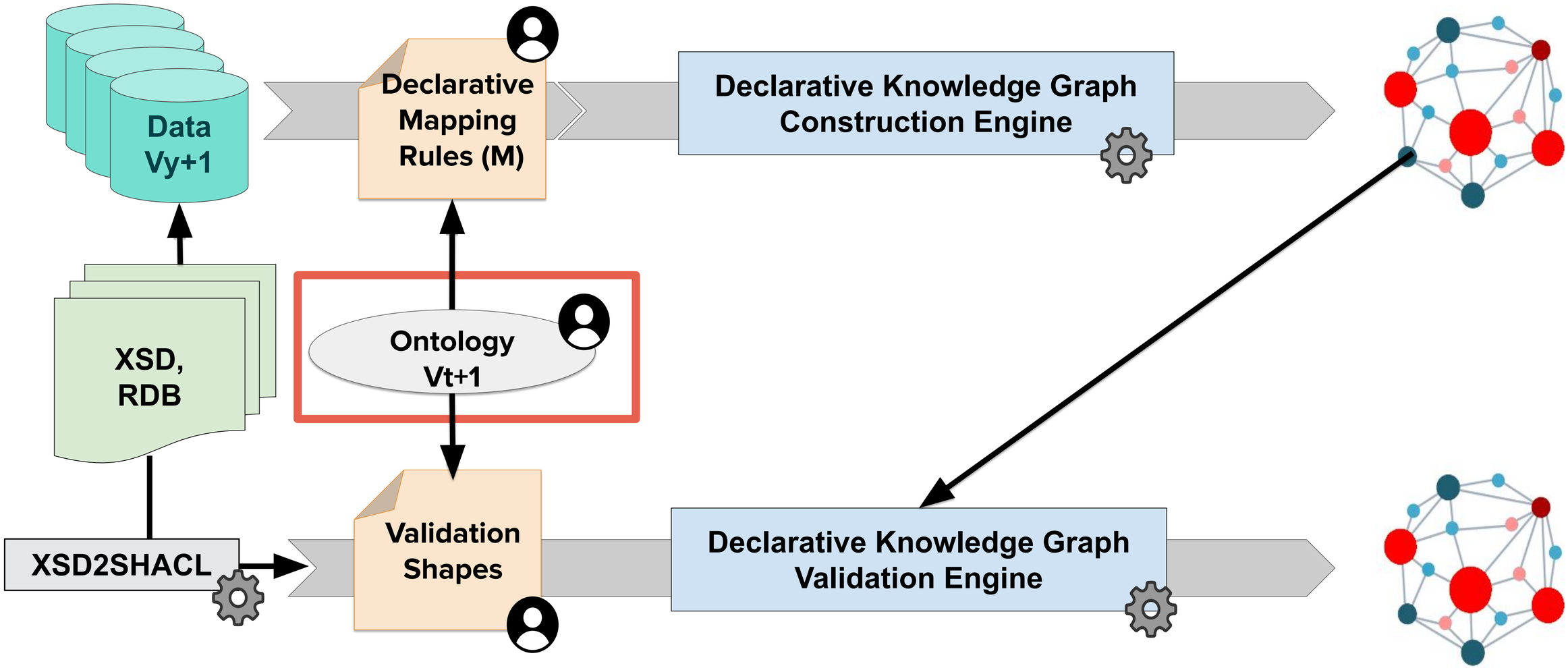
- data source changes?

- metadata representation model?

- **the ontology changes?**

DISCLAIMER:
THIS IS ONGOING WORK, FEEDBACK IS APPRECIATED

Ontology changes impact over the KGs

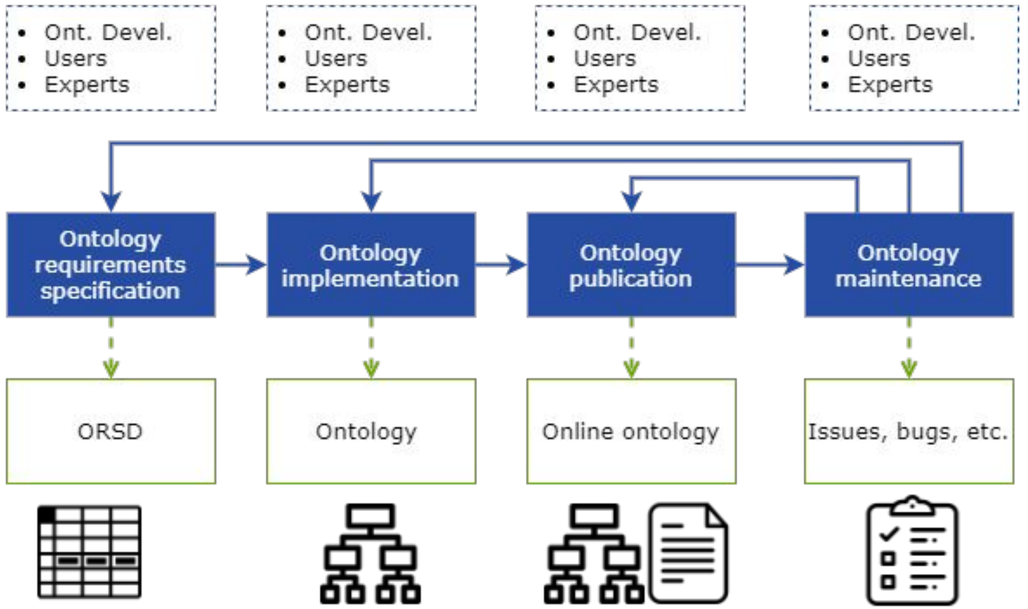


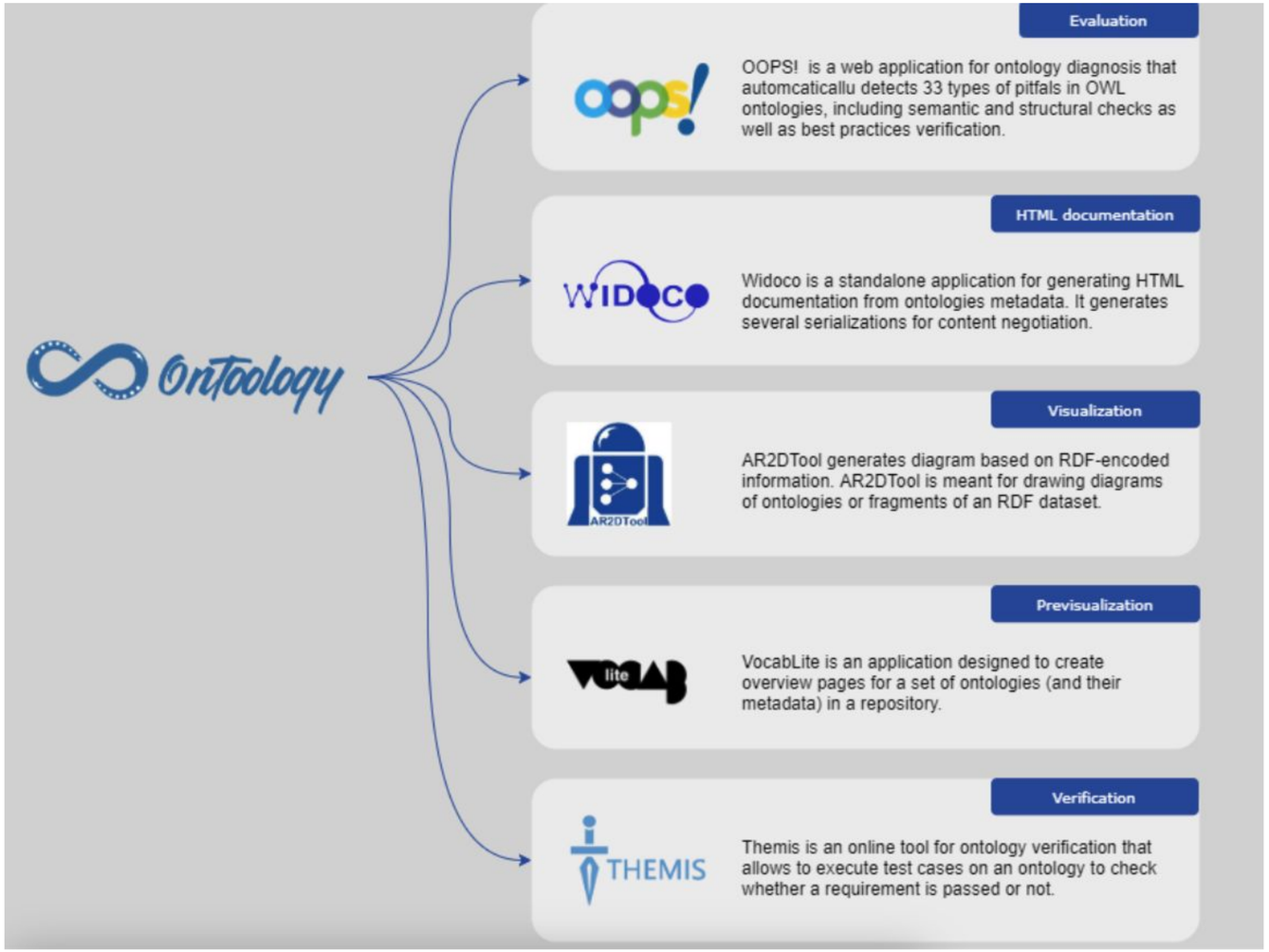
Changes in e-Procurement Ontology

class	added property	deleted property
epo:AwardDecision	generalisation → epo:ProcurementElement	generalisation → epo:ProcurementObject
epo:Contract	epo:signedByContractor → epo:Contractor	epo:signedBySignatory → epo:ContractSignatory
epo:Contract	generalisation → epo:ProcurementElement	generalisation → epo:ProcurementObject
epo:Contract	epo:signedByBuyer → epo:Buyer	
epo:ProcurementObject	epo:hasPurpose → epo:Purpose (moved from epo:Lot)	
epo:ProcurementObject	epo:usesChannel → cv:Channel (moved from epo:PlannedProcurementPart)	
epo:ProcurementObject	epo:refersToPlannedPart → epo:PlannedProcurementPart (moved from epo:Procedure)	
epo:ProcurementObject	epo:hasEstimatedValue → epo:MonetaryValue (epo:hasEstimatedValue:epo:MonetaryValue	

https://docs.ted.europa.eu/EPO/latest/release-notes.html

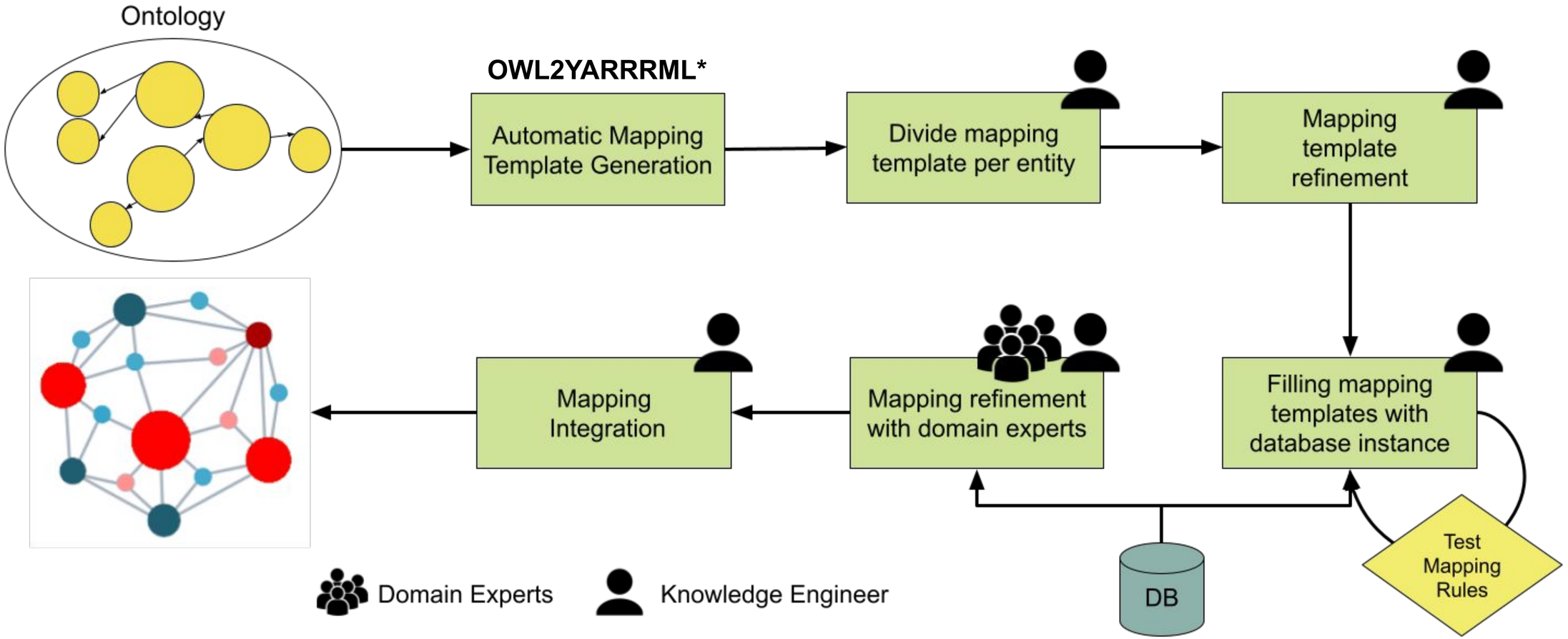
LOT: Linked Open Terms Methodology





Poveda-Villalón, M., Fernández-Izquierdo, A., Fernández-López, M., & García-Castro, R. (2022). LOT: An industrial oriented ontology engineering framework. Engineering Applications of Artificial Intelligence, 111, 104755.

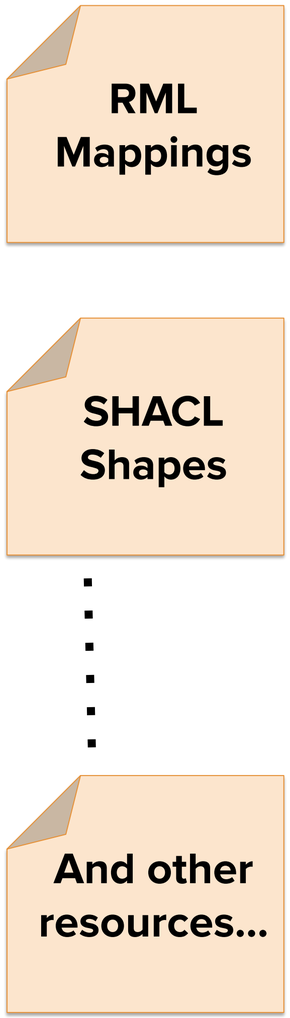
First naïve approach...



* https://github.com/oeg-upm/owl2yarrrml

Chaves-Fraga, D., Corcho, O., Yedro, F., Moreno, R., Olías, J., & De La Azuela, A. (2022). Systematic construction of knowledge graphs for research-performing organizations. *Information, 13(12), 562.*

LOT4KG?



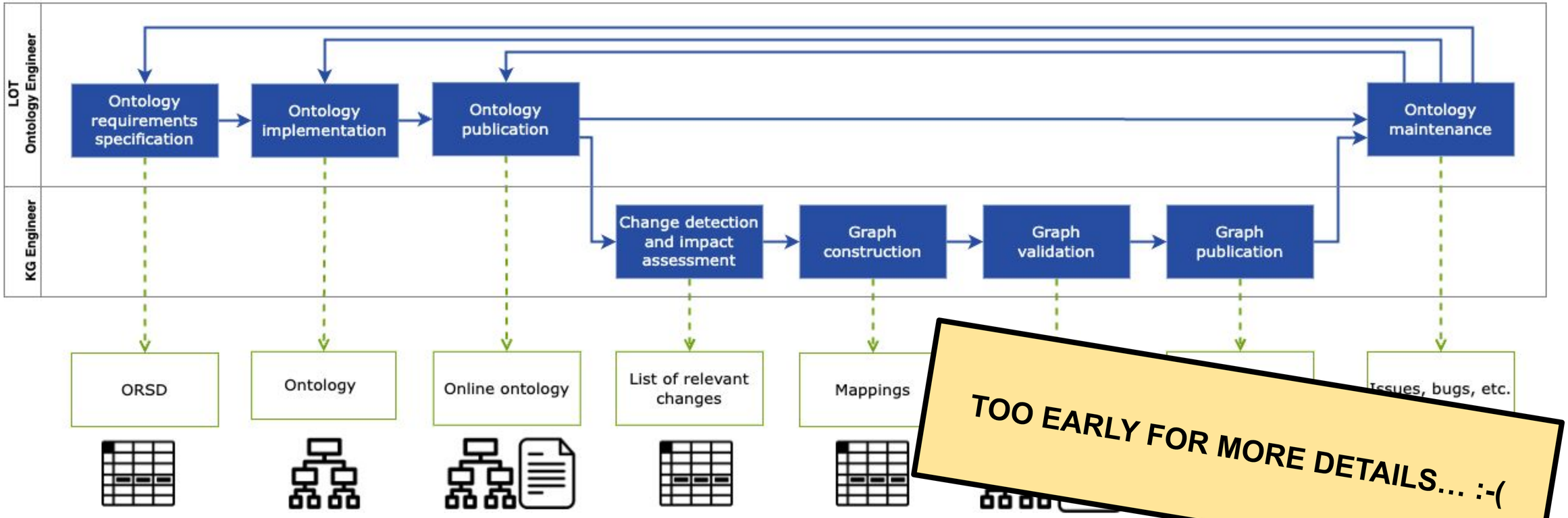
Documentation
- Mapping templates? https://github.com/oeg-upm/owl2yarrrml?
- Mapping/Shape patterns?
- How mappings are related with the ontology terms?

Validation
- Are the mapping syntactically/semantically correct?
- Provenance of data constraints?

Evaluation
- Do we have mapping or shape pitfalls??
- F.A.I.R. mapping/shapes?

Visualization
- Human-friendly visualizations
- Mapping/shapes changes/versioning

LOT4KG: An ontology-driven methodology for KG Construction and Validation



- Re-use an existing ontology for describing the changes between two versions of the same ontology
- Explicitly define the **impact of the changes on mappings and shapes**
- Automatic propagation of changes → Mapping and shape versioning (benefits of RDF)

My two cents on Data Management 4 KG Ecosystems

Still room for performance improvements in KG Construction and Validation **but…**

1) We need to pay more attention to other aspects in our projects (e.g., maintainability)
2) Documentation, evaluation, etc. of related assets to the KGs needs to be defined
3) Everything is related and we should exploit that relationships (automation?)
4) Ontologies are the backbone for an intelligent KG management
5) Evolution (i.e., changes) is a parameter we must consider, increasing the complexity of our systems in all challenges (scalability, maintainability, etc)
6) There is still too much to do for supporting on production real-projects

Next steps: Dagstuhl Seminar

Are Knowledge Graphs Ready for the Real World? Challenges and Perspective

Long-Large Dagstuhl Seminar (4th to 10th February 2024)

David
Chaves-Fraga

Oscar
Corcho

Anastasia
Dimou

Maria-Esther
Vidal

Challenges on Data Management for Evolving Knowledge Graphs

David Chaves-Fraga
CiTIUS@University of Santiago de Compostela (Spain)
david.chaves@usc.es

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

CAMPUS VIDA CAMPUS DE EXCELENCIA INTERNACIONAL

CiTIUS Singular Research Center on **Intelligent technologies**