« I want a KG-based Digital Twin, and that's all I (want to) know »

or from Technology Readiness Level 4 to above with Knowledge Engineering and Data Model Operations?

Lionel Tailhardat, Orange

lionel.tailhardat@orange.com

https://genears.github.io/

KG4DI workshop - 11 Dec, 2024

orange™

Knowledge Graphs for Networks Operations?

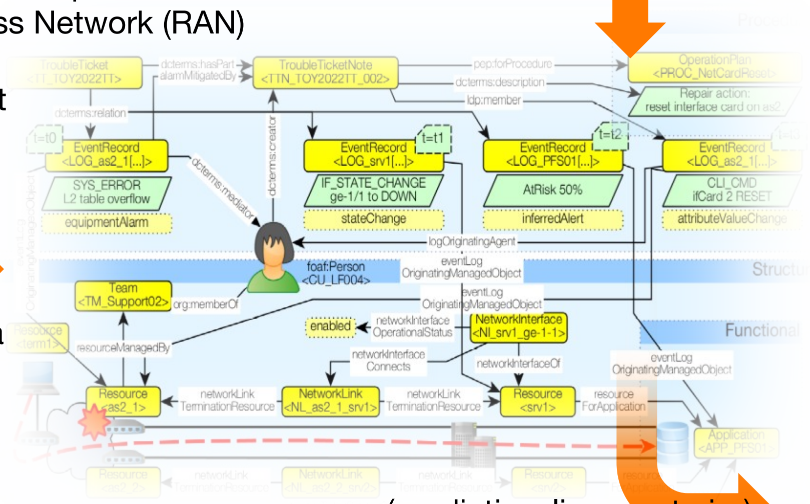



Multiple networks, multiple perspectives ...

Services
5G core
IP backbone
Long-haul / transport network
Radio Access Network (RAN)
IT/PFS
Environment

Building a Unified View

Cross-Network User and Equipment Behavioral Analysis



(predictive, live, posterior)

Incident Management
- Root Cause Analysis
- Alarm Correlation & Enrichment
- Incident Impact Analysis

Change Management
- Change Impact Analysis
- Preventing Concurrent Operations

Quality of Service Monitoring
- Network Performance and Availability
- Consumer and Business Client Dashboard

Automation and Tools
- Automatic Asset Configuration
- Network Assets & Flows Discovery
- Data Quality & Consistency of Reference Databases

L. Tailhardat, et al. **NORIA - Machine LearNing, Ontology and Reasoning for the Identification of Anomalies.** Position poster, IA2 – SCAI, 2021.
Y. Chabot, et al. **NORIA: Network anomaly detection using knowledge graphs.** Blog article in Orange – Hello Future, 2024.
L. Tailhardat, et al. **Knowledge Graphs for Enhanced Cross-Operator Incident Management and Network Design**. IETF Internet Draft, 2024.
M. Mackey, et al. **Knowledge Graph Framework for Network Operations**. IETF Internet Draft, 2024.

Research Questions

How to define an anomaly model in a dynamic technical environment with various interdependencies, and what form should this model take to be shareable among practitioners and directly usable in anomaly detection tools and decision support systems?

RQ. 1

Anomaly model production & utilization with heterogeneous data
What is an adequate neuro-symbolic AI architecture that can learn logically-constrained behavioral rules from events and topology data of an ICT system, and enable to detect and interpret complex anomalous technical or user-based situations?

RQ. 2

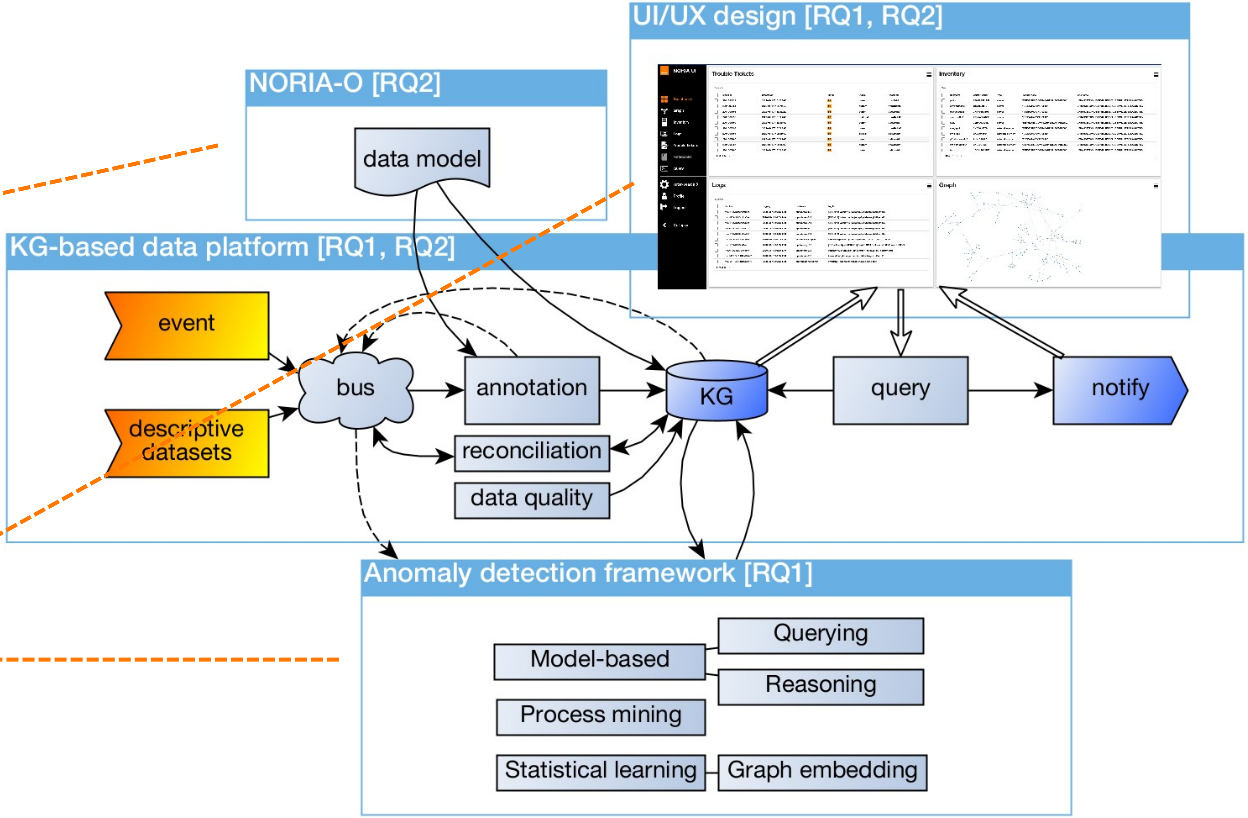
Constraints on the internal representation of data and knowledge
Can human operators and decision support AI agents use the same Knowledge Representation (KR) of ICT systems for anomaly detection and knowledge management, that KR being subject to computation efficiency and interpretability?

L. Tailhardat. **Anomaly Detection using Knowledge Graphs and Synergistic Reasoning: Application to Network Management and Cyber Security.** PhD thesis, EURECOM, 2024.

Research Roadmap

Building a **graph** for dynamic ICT systems

Exploiting the ICT systems **knowledge**

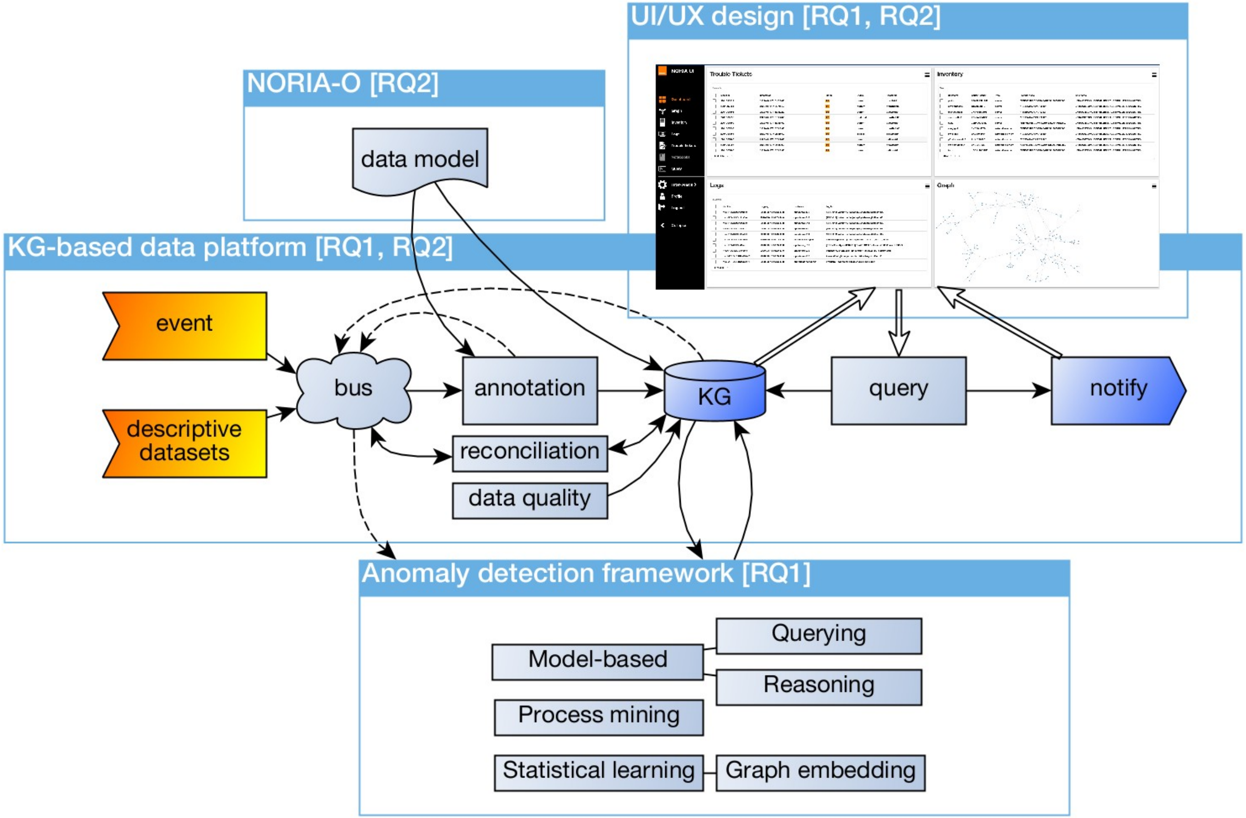


RQ. 1: Anomaly model production & utilization with heterogeneous data
RQ. 2: Constraints on the internal representation of data and knowledge
ICT: Information & Communications Technology

L. Tailhardat. **Anomaly Detection using Knowledge Graphs and Synergistic Reasoning: Application to Network Management and Cyber Security.** PhD thesis, EURECOM, 2024.

Research Summary

- **Holistic perspective** on the application domain.
- **Explicit representation** of networks and their ecosystem.
- Algorithmic techniques heavily reliant on **formal representation** at the level of generated models or their results.



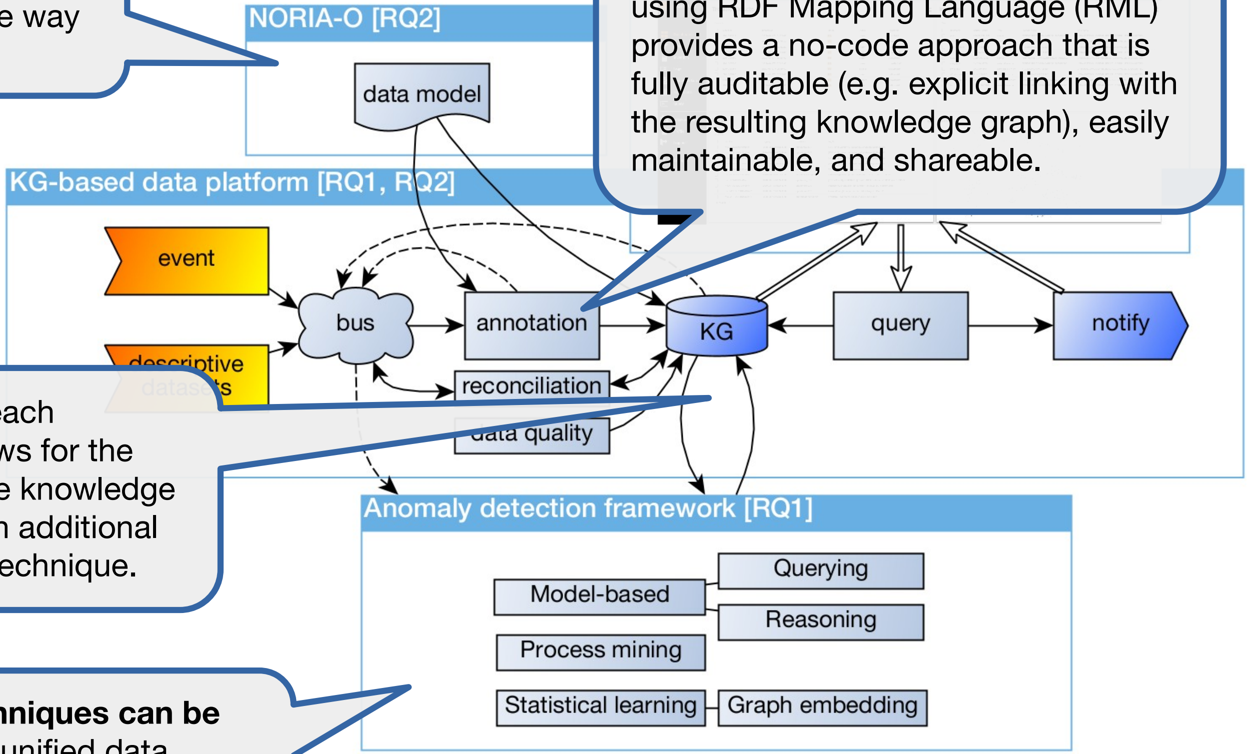
Now in position to :

- Achieve **cross technical domain anomaly detection** with intrinsic explainability and probabilistic reasoning capabilities.
- Identify and share strengths and weaknesses of infrastructures (FMEA).

Research Summary

- ✓ **Explicit representation of** networks and their ecosystem
- ✓ **Declarative data transformation** using RML, patching queries, and generic KGC pipelines.
- ✓ **Opensource and Semantic Web protocol stack**, fostering the adoption of the knowledge graph paradigm at scale by the NetOps & SecOps communities.

Ontologies bring **unified view of heterogeneous systems**, including their dynamics, in line with the way experts refer to their network.



Declarative data transformation: using RDF Mapping Language (RML) provides a no-code approach that is fully auditable (e.g. explicit linking with the resulting knowledge graph), easily maintainable, and shareable.

Cooperative decision-making: each technique, taken individually, allows for the **reinjection of knowledge** into the knowledge graph, which can then serve as an additional contextual element for a second technique.

Now in position to:

- Achieve **cross technical domain anomaly detection** with intrinsic explainability and probabilistic reasoning capabilities.
- Identify and share strengths and weaknesses of infrastructures (FMEA).

Anomaly detection techniques can be more generic thanks to unified data representation, rather than being specialized in a specific technical domain.

Large-scale deployment with KE-Ops & DM-Ops?

Perspectives

Presenting the previous solutions (Technology Readiness Level 4/5/6) to operation experts inevitably raises questions related to their integration and deployment (Technology Readiness Level 7/8/9) ...

I want (my) KG-based Digital Twin

How can I assess the impact of **changes in data sources or ontology** modifications?

What **new skills** are needed to deploy KG-based solutions and manage system coherence and lifecycle, and how will this impact the current organization?

How can we efficiently design, implement, and manage **overlapping use cases** from different technical domain owners?

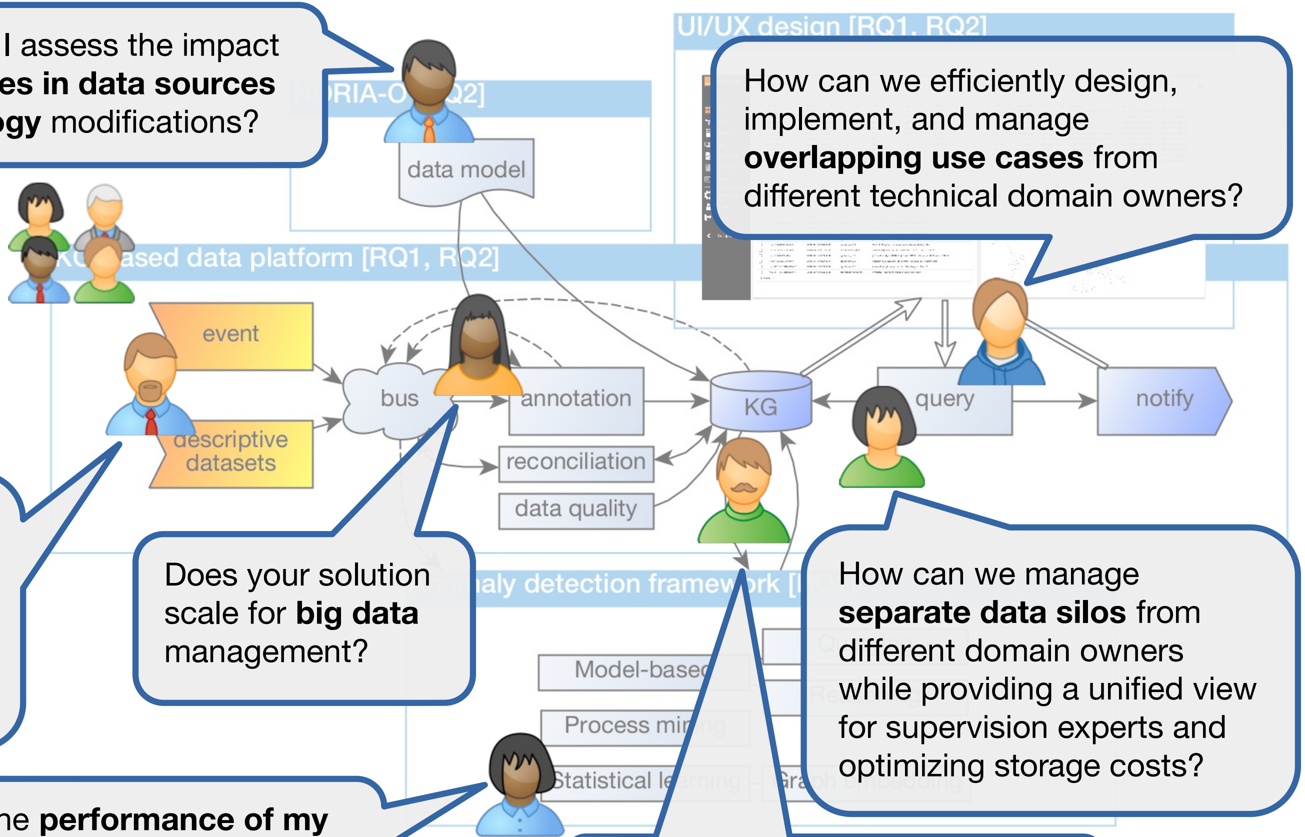
What tools can assist in implementing and **managing mapping rules**, given the challenges of time consumption, required skills (modeling, abstraction, data integration), and the need for domain expertise?

Does your solution scale for **big data** management?

How can we manage **separate data silos** from different domain owners while providing a unified view for supervision experts and optimizing storage costs?

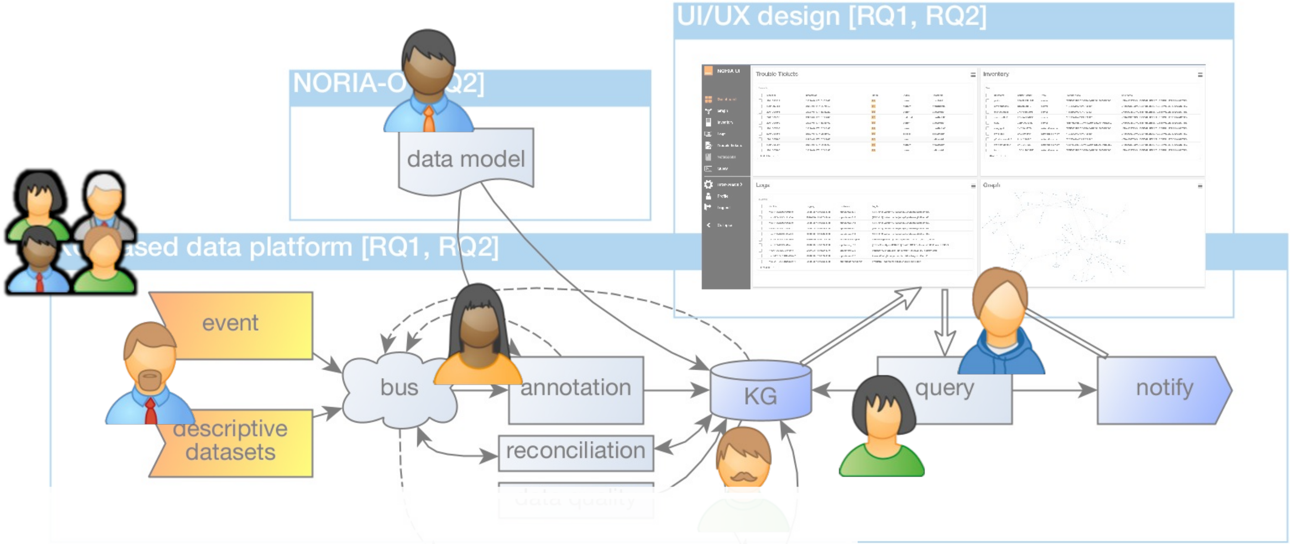
How can I assess the **performance of my detection algorithms** for rare events without all the necessary testing data?

How can I avoid **ever-expanding knowledge graphs** ?



I want (my) KG-based Digital Twin

What **new skills** are needed to deploy KG-based solutions and manage system coherence and lifecycle, and how will this impact the current organization?



Challenges

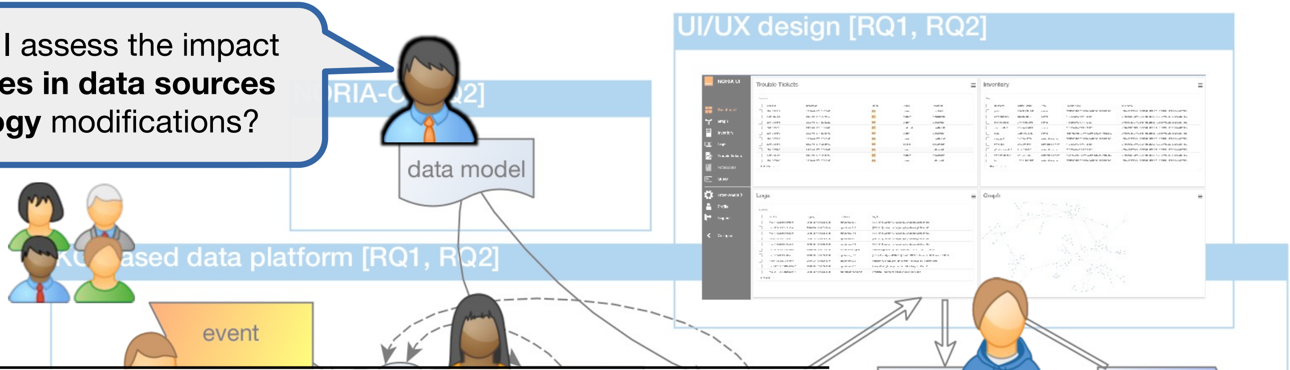
- ✗ Need to **understand the entire technological stack** of knowledge graphs and the Semantic Web,
- ✗ **Few blueprints for large systems** or organizations to guide role and responsibility breakdowns.

Guidance (partial list)

- ➢ Katariina Kari. « What Over **7 Years of Building Enterprise Knowledge Graphs** Has Taught Me About Theory and Practice », ESWC, 2024.
- ➢ « Introduction to a Web of Linked Data », INRIA @ fun-mooc.fr
- ➢ B. Steenwinckel, et al. « FLAGS: A Methodology for Adaptive Anomaly Detection and Root Cause Analysis on Sensor Data Streams by Fusing Expert Knowledge with Machine Learning ». FGCS, 2021.

I want (my) KG-based Digital Twin

How can I assess the impact of **changes in data sources or ontology** modifications?



Challenges

✗ Need to relate **data governance** practices with **knowledge engineering** practices,

✗ Be able to **plan and organize changes** so that they are reversible and assessable (A/B testing).

Guidance (partial list)

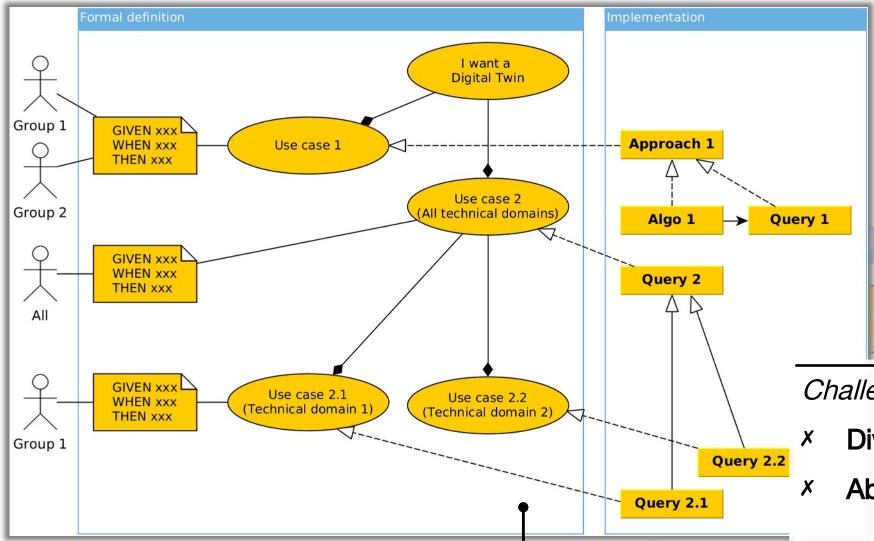
➢ D. C. Herreros, et al. « **Propagating Ontology Changes** to Declarative Mappings in Construction of Knowledge Graphs », KGCW, 2024.

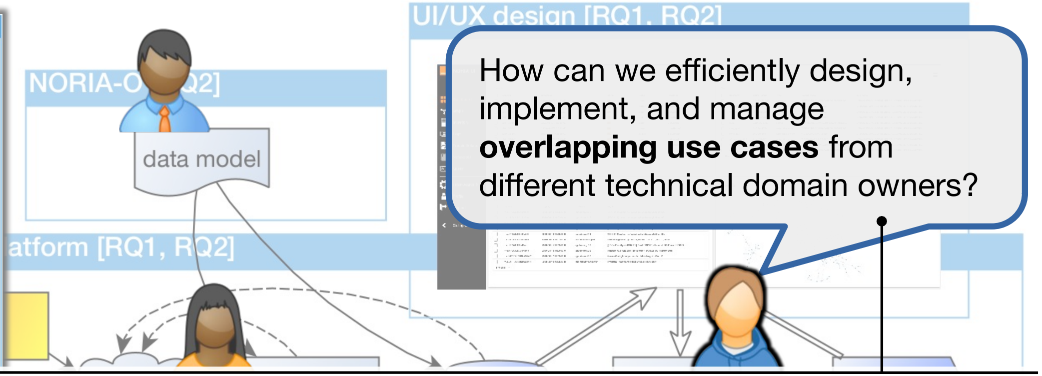
➢ Tools and techniques from « SemTab: Semantic Web Challenge on Tabular Data to Knowledge Graph Matching », AIDA & SIRIUS & IBM.

➢ M. Lenzerini. « **Data integration: A theoretical perspective** », PODS, 2002.

➢ « Abstract Reduction System (ARS) » theory and tools.

I want (my) KG-based Digital Twin





Challenges

✗ **Diverse terminology for similar use cases** across technical domains,

✗ **Abstraction skills** are not central to NetOps and SecOps practices.

Guidance (partial list)

➢ **Design theory** and best practices
- Tools and practices from the « Concept-Knowledge theory (C-K theory) ».
- Alistair Cockburn. « Writing Effective Use Cases », Addison-Wesley, 2012.
- SAFe – Story. https://scaledagileframework.com/story/. Scaled Agile, Inc. 2022.
- ISO/IEC/IEEE International Standard - « Systems and software engineering – Life cycle processes – Requirements engineering ». IEEE, 2018.
- « Gherkin language », cucumber.io

➢ Organize the **graph exploitation techniques** hierarchically through **factorization**
- H. Knublauch. « DASH SPARQL Templates Vocabulary », 2021.

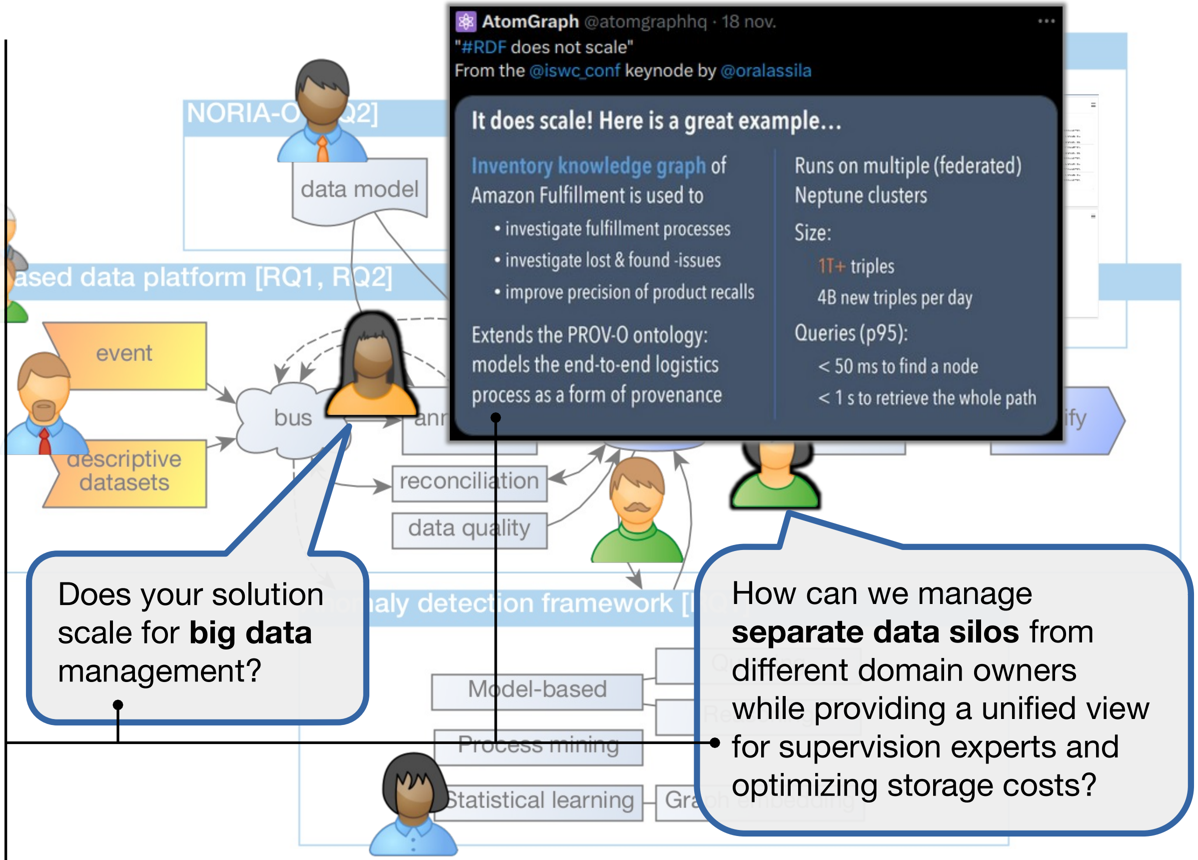
I want (my) KG-based Digital Twin

Challenges

- ✗ **High-paced & high-volume data**, e.g. national transmission network (WDM/SDH/microwave):
  - Assets: 100M nodes & node-node relations,
  - Alarming: 2M events/day, with occasional 5K events/minute bursts.,
- ✗ Many SGBD vendors with proprietary data models, local analytics frameworks, and partial/flavored implementation of representation/query standards (**vendor lock-in**).

Guidance (partial list)

- **Hybrid** local/hyperscaler graph/SQL/no-SQL stores & analytics **architectures** + **stream data summarization** at the knowledge graph construction step.
  - « SPARQL 1.1 Federated Query », W3C, 2013.
  - O. Lassila, et al. « The OneGraph vision: Challenges of breaking the graph model lock-in ». Semantic Web, 2022.
  - L. Tailhardat, et al. « Knowledge Graphs for Enhanced Cross-Operator Incident Management and Network Design”. IETF Internet Draft, 2024.



I want (my) KG-based Digital Twin

Challenges

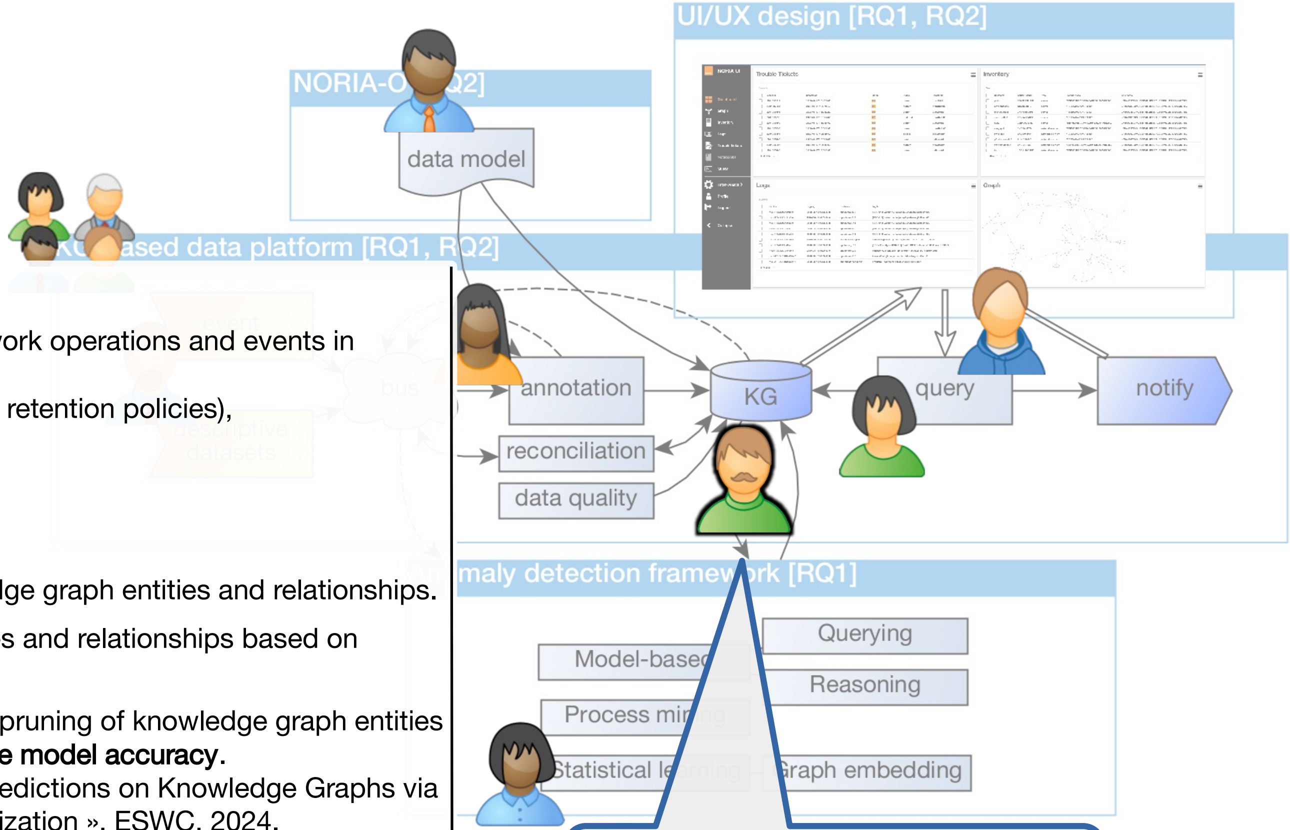
✗ Handling **long duration storage** of the network operations and events in accordance to:
  - Legal and business requirements (data retention policies),
  - ML/DL model training requirements.

Guidance (partial list)

➢ Add a **time-to-live** (TTL) tag to the knowledge graph entities and relationships.

➢ Periodically **prune knowledge graph** entities and relationships based on provenance annotations.

➢ Perform **graph summarization** or selective pruning of knowledge graph entities and relationships, ensuring **stable inference model accuracy**.
  - R. Barile, et al. « Explanation of Link Predictions on Knowledge Graphs via Levelwise Filtering and Graph Summarization », ESWC, 2024.
  - S. H. Hahn, et al. « RDF2vec Embeddings for Updateable Knowledge Graphs – Reuse, don't Retrain! », ESWC, 2024.



How can I avoid **ever-expanding knowledge graphs** ?

I want (my) KG-based Digital Twin

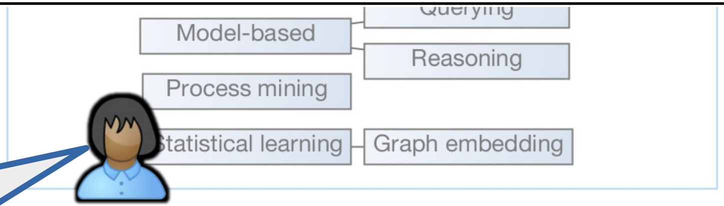
Challenges

- ✗ **Sharing dataset** (e.g. network topology, failure modes, remediation procedures) for testing purposes **without revealing business data.**
- ✗ **Ontologies** enable represention and reasoning within a specific application domain, but they **are not the use case itself** (e.g. events on a network topology versus detecting a resilience issue based on these events).

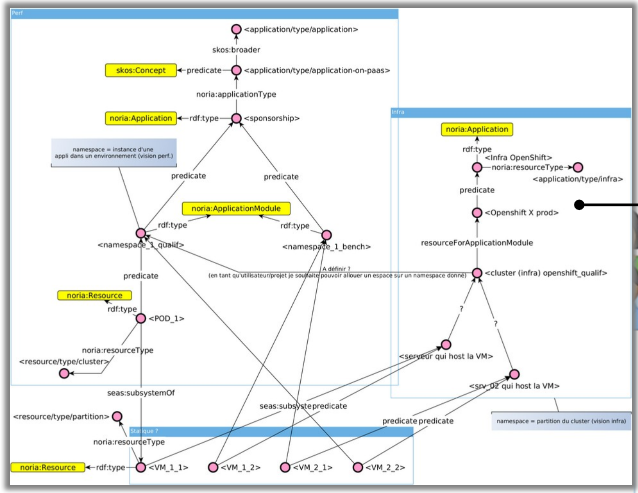
Guidance (partial list)

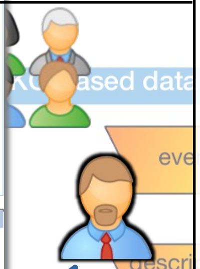
- **Generating knowledge graphs**
  - N. Hubert, et al. « PyGraft: Configurable Generation of Synthetic Schemas and Knowledge Graphs at Your Fingertips », ESWC, 2024.
  - M. Vecovska, et al. « RDFGraphGen: A Synthetic RDF Graph Generator based on SHACL Constraints ». arXiv, 2024.
- Sharing **failure modes** and **situation descriptions** in a **standardized form**
  - B. Steenwinckel, et al. « Towards Adaptive Anomaly Detection and Root Cause Analysis by Automated Extraction of Knowledge from Risk Analyses », ISWC, 2018.
  - V. Riccobene, et al. « Experiment: Network Anomaly Lifecycle », IETF Internet Draft, 2024.

How can I assess the **performance of my detection algorithms** for rare events without all the necessary testing data?



I want (my) KG-based Digital Twin





What tools can assist in implementing and **managing mapping rules**, given the challenges of time consumption, required skills (modeling, abstraction, data integration), and the need for domain expertise?

Challenges

✗ Testing **mapping hypothesis** and ensuring **mapping alignment**:
- Technical domains can all leverage a same (set of) ontologies,
- But all requires specific ways of combining and mapping data to concepts and relationships to reflect the networks characteristics (e.g. a flat meshed IP core network versus a stacked transmission network versus a Kubernetes CaaS system).

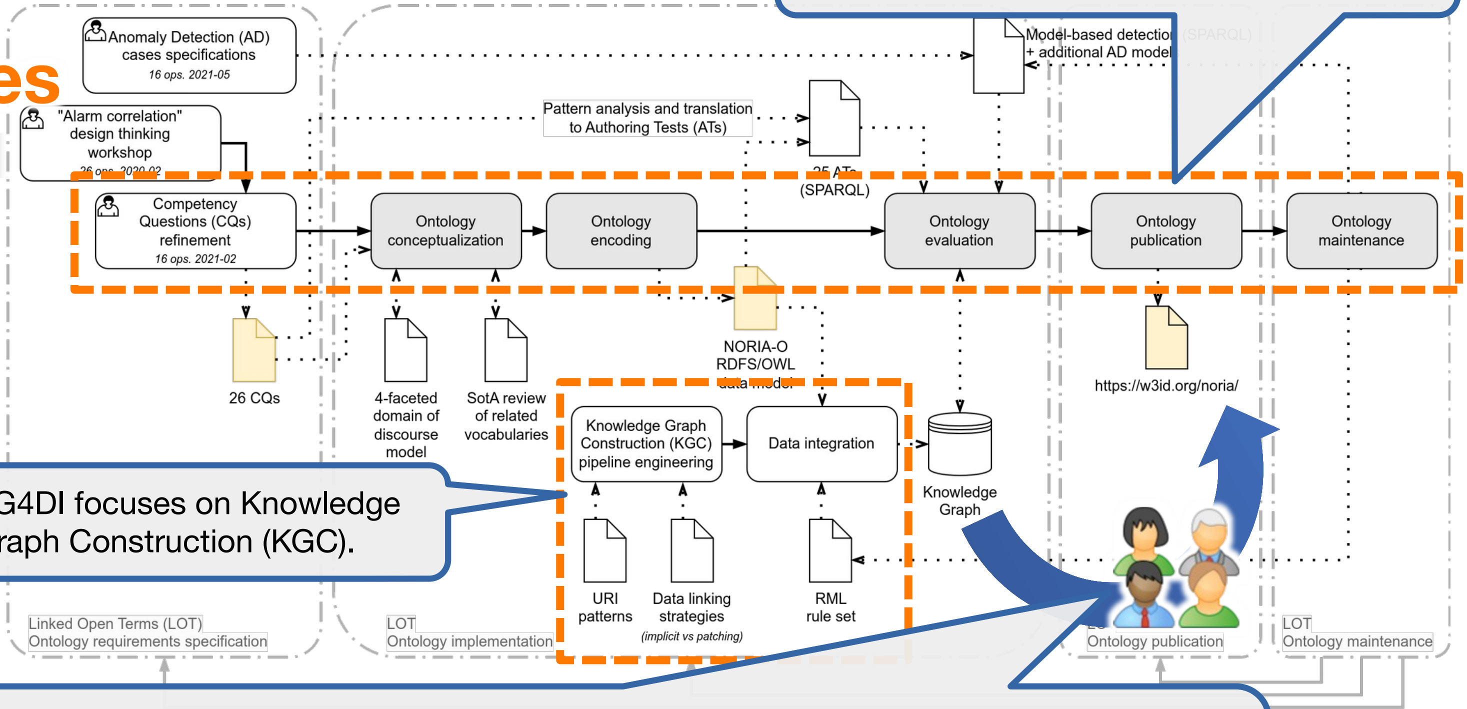
Guidance (partial list)

- An existing set of **tools to integrate in a coherent workflow**
  - RMLEditor (https://app.rml.io/rmleditor/), YARRRML (https://rml.io/yarrrml/), Matey (https://rml.io/yarrrml/matey/).
  - J. Toledo, et al. « RML mapping documentation », 2024.
  - P. Colpaert, et al. « TurtleValidator », IDLab, 2014.
  - « Shapes Constraint Language (SHACL) », W3C, 2017.
  - A. Dimou, et al. «Assessing and Refining Mappings to RDF to Improve Dataset Quality », 2015.
- Tools and techniques from « SemTab: Semantic Web Challenge on Tabular Data to Knowledge Graph Matching », AIDA & SIRIUS & IBM.
- **Reusing network specific data models**
  - L. Tailhardat, et al. « Knowledge Graphs for Enhanced Cross-Operator Incident Management and Network Design”. IETF Internet Draft, 2024.
  - M. Mackey, et al. « YANG to RDF », IETF 121 meeting, 2024.

Shifting **Perspectives**

Conclusion

Knowledge Engineering (KE) methodologies (e.g. LOT) focus on ontologies.



KG4DI focuses on Knowledge Graph Construction (KGC).

Ops look at ontologies through the lenses of the KGC stage

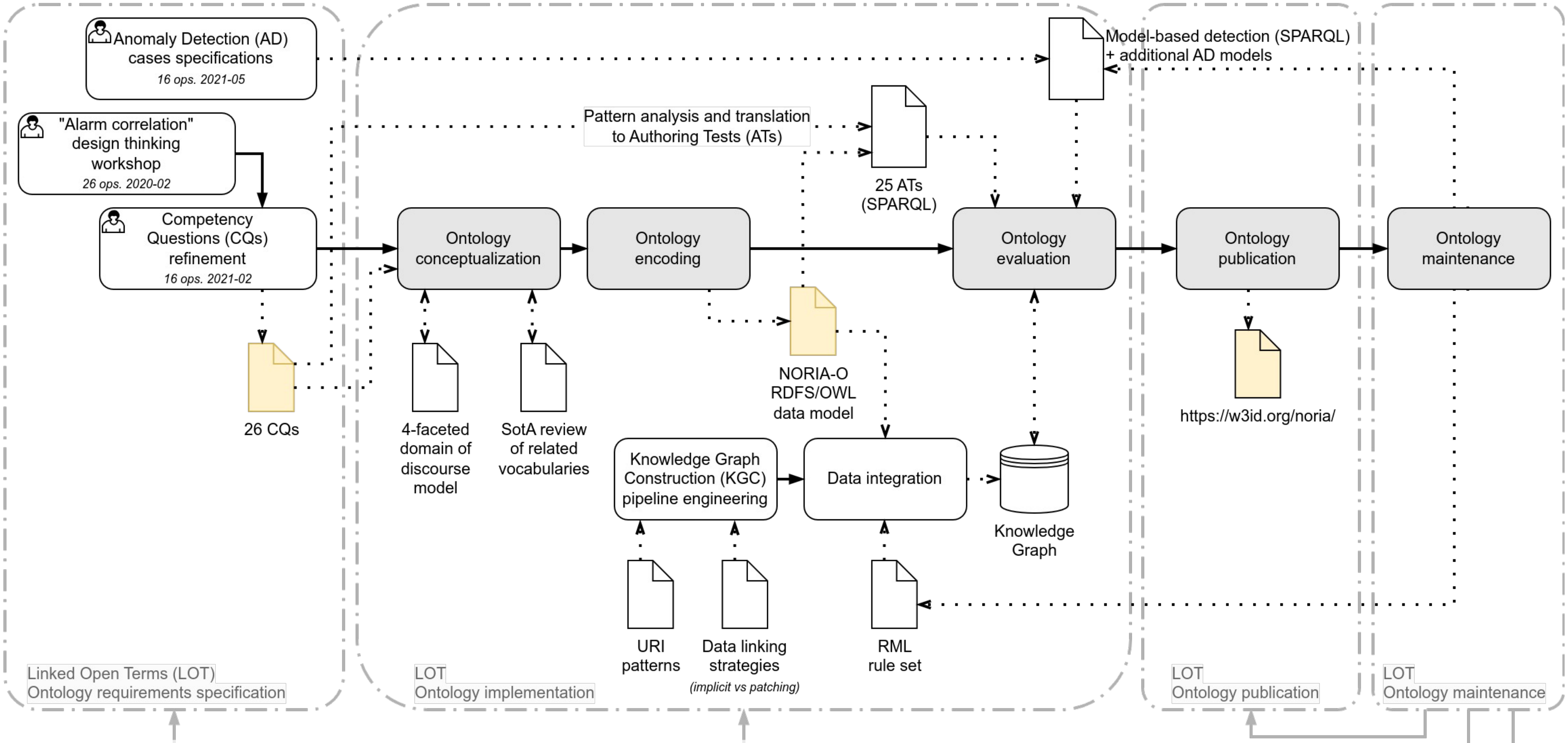
- **KE-Ops / DM-Ops**? i.e. extend KE methodologies with KGC to bring an end-to-end guidance.
- **Gather KG-based system blueprints / proposals** in a standardized format for research purposes? e.g. SweMLS-KG (F. J. Ekaputra et al. 2023) + Semantic Web Tool Hub (A. Reiz, et al. 2024).

L. Tailhardat. **Anomaly Detection using Knowledge Graphs and Synergistic Reasoning: Application to Network Management and Cyber Security.** PhD thesis, EURECOM, 2024.

Additional materials

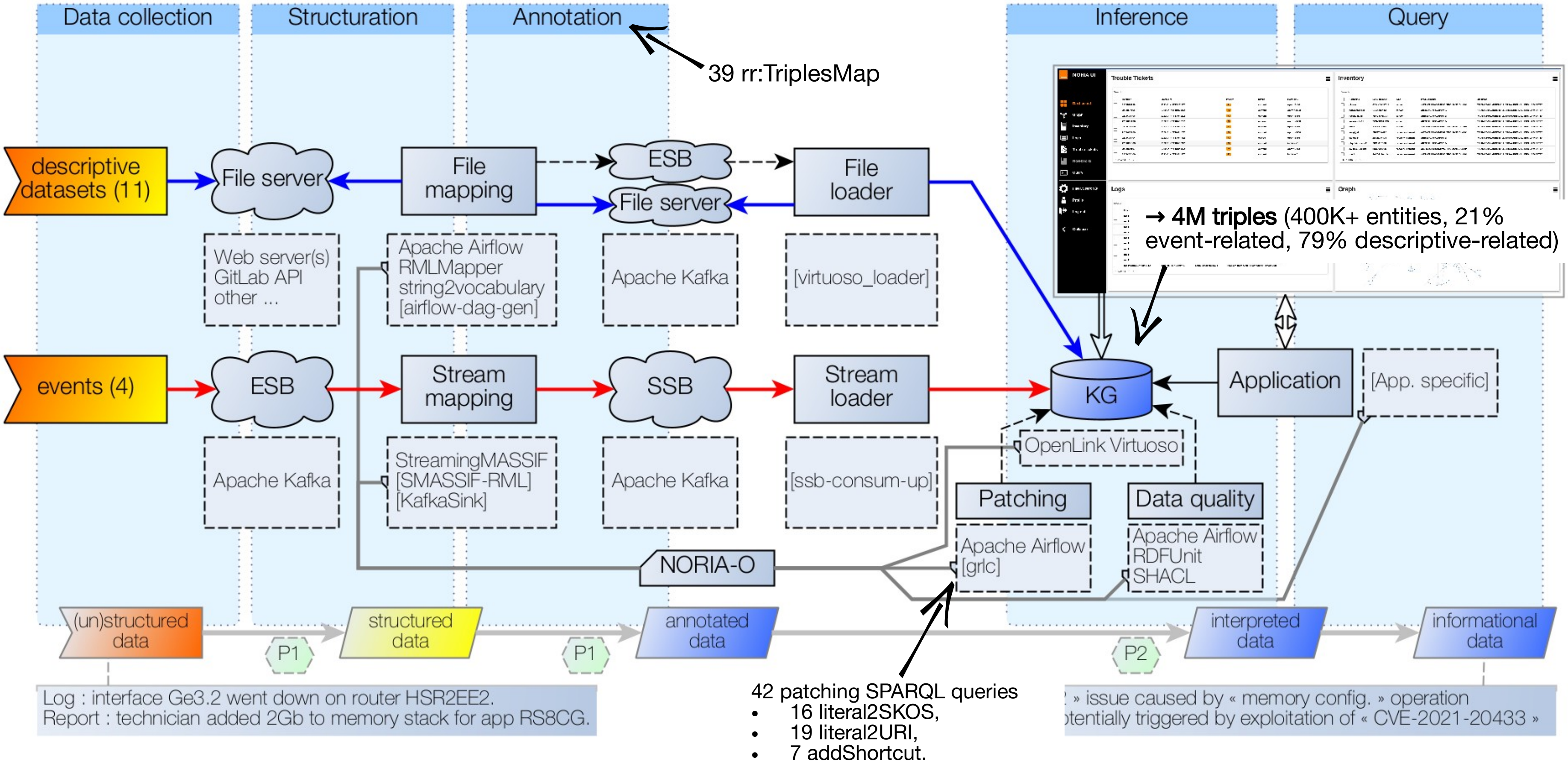
Appendix

Knowledge Engineering



L. Tailhardat. **Anomaly Detection using Knowledge Graphs and Synergistic Reasoning: Application to Network Management and Cyber Security.** PhD thesis, EURECOM, 2024.

Knowledge Graph Construction

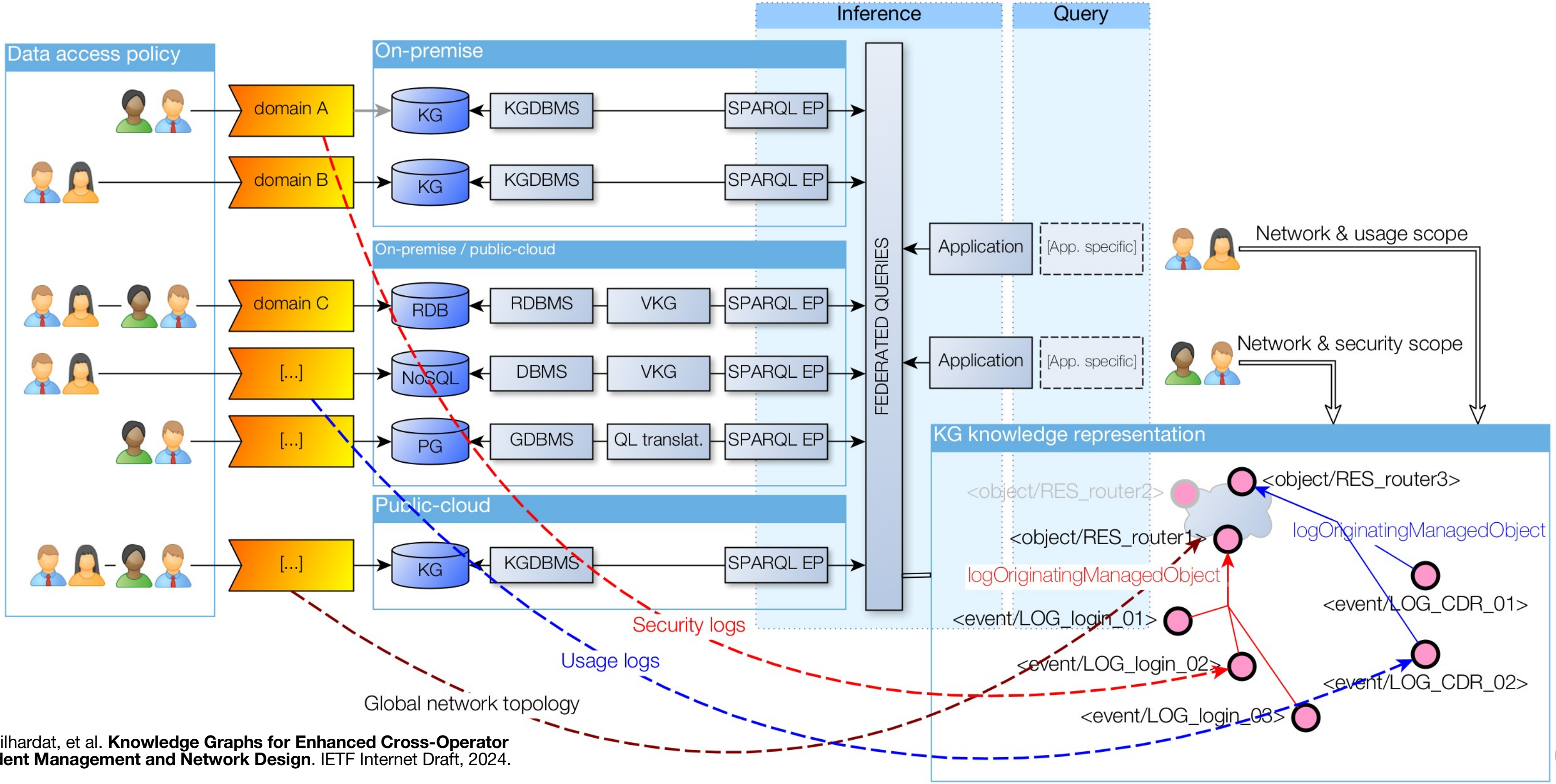


L. Tailhardat et al. **Designing NORIA: a Knowledge Graph-based Platform for Anomaly Detection and Incident Management in ICT Systems**. ESWC'23.

Federating Partitioned Data

Federated queries for providing,

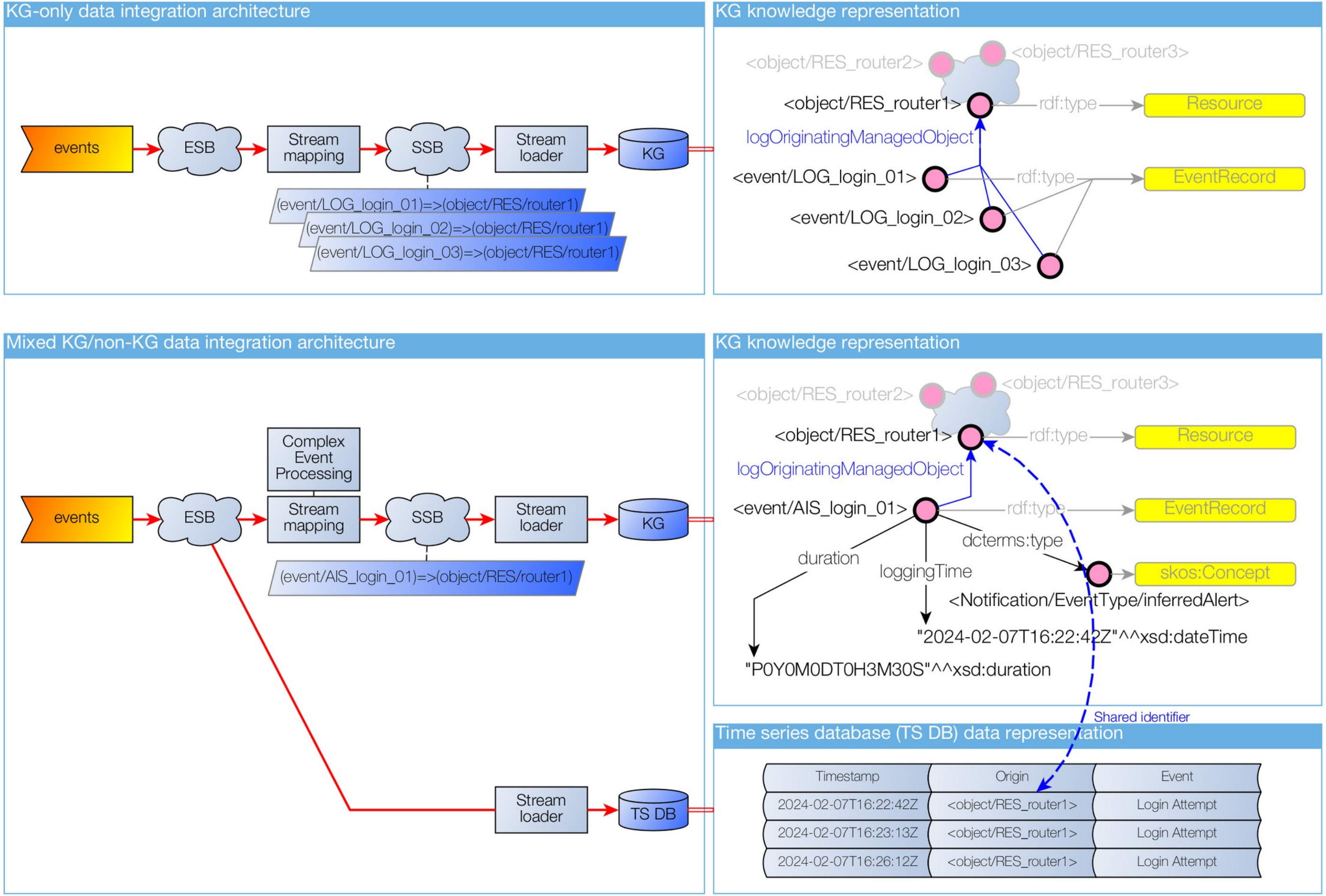
- A single protocol to access data silos using different storage technologies & formalisms,
- A unified representation of data domains with scoped access control.



L. Tailhardat, et al. **Knowledge Graphs for Enhanced Cross-Operator Incident Management and Network Design**. IETF Internet Draft, 2024.

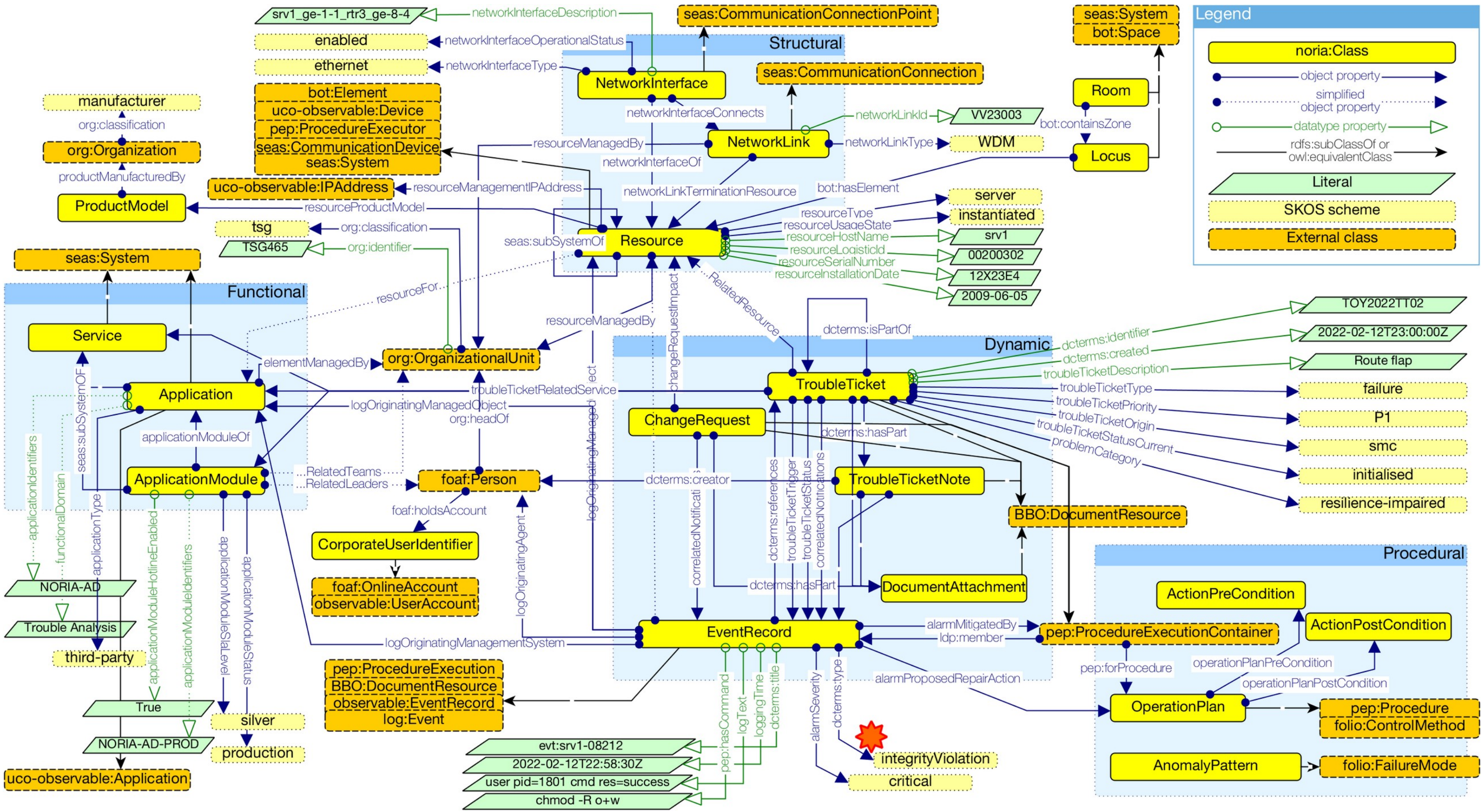
Scaling with Streams

- Building the graph with **all incoming data**.
- Building the graph with **summarized data**, and ensure **unicity of object identifers** across data stores.



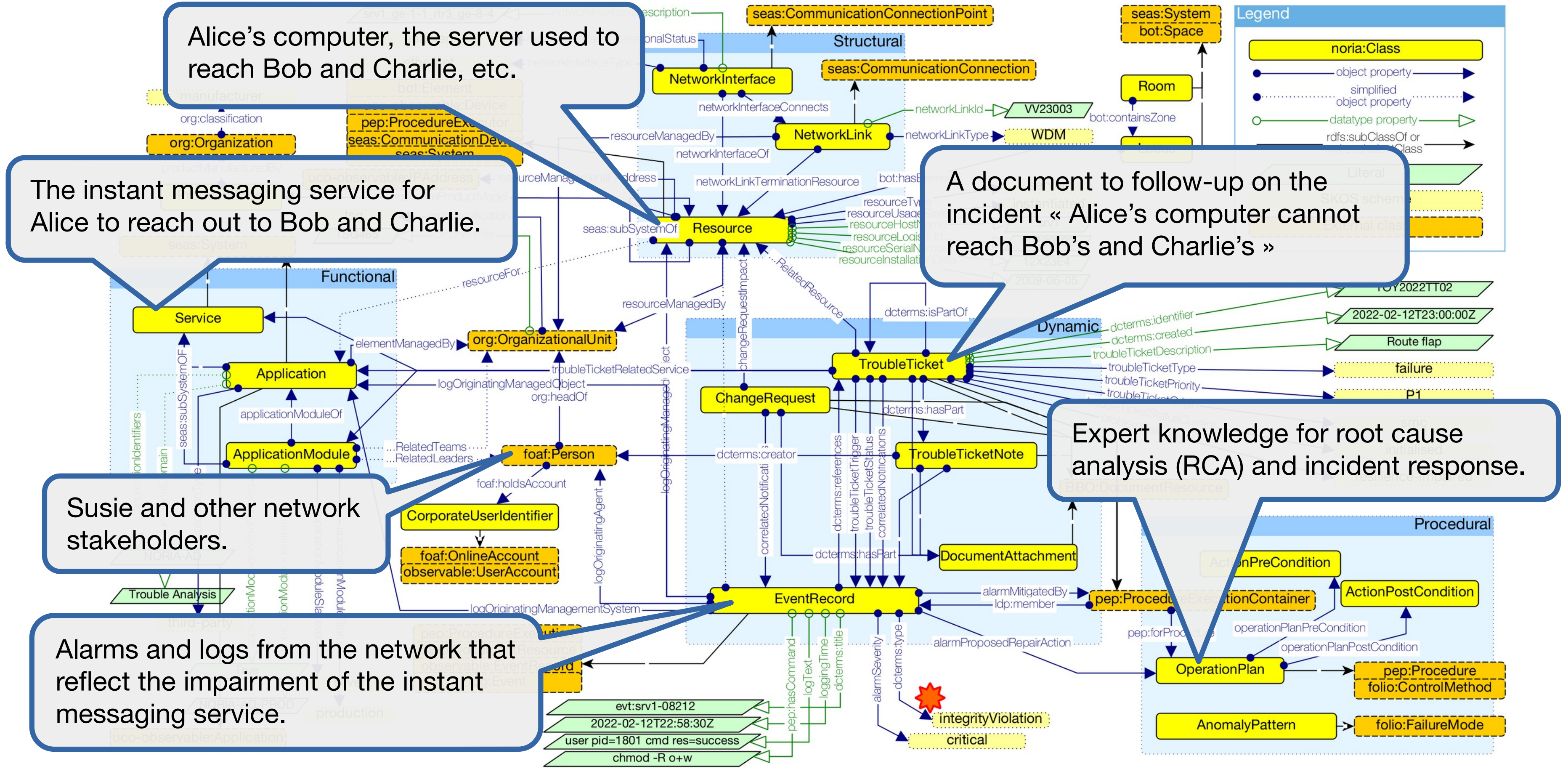
L. Tailhardat, et al. **Knowledge Graphs for Enhanced Cross-Operator Incident Management and Network Design**. IETF Internet Draft, 2024.

An ontology for Dynamic ICT systems



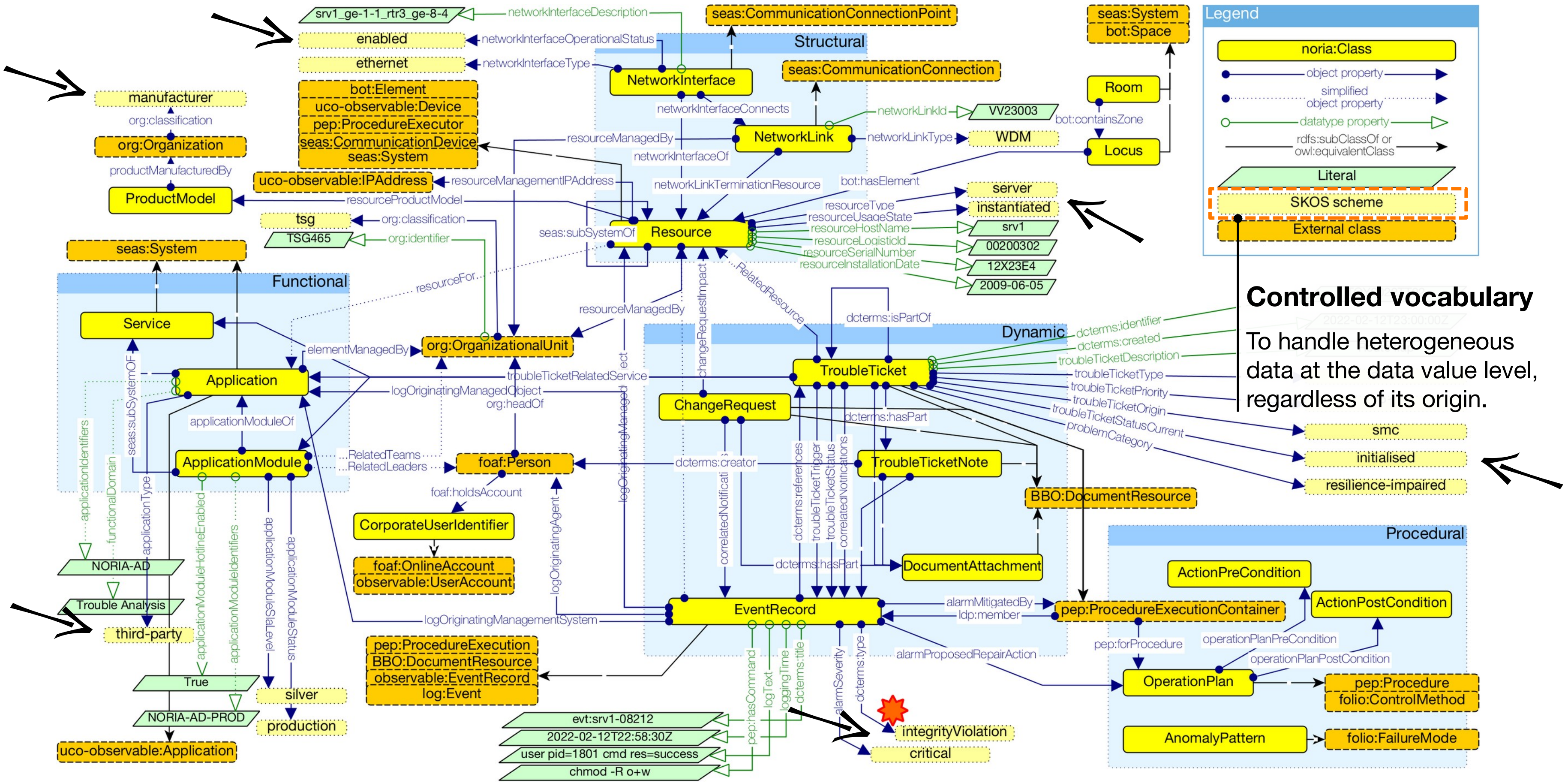
L. Tailhardat et al. **NORIA-O: An Ontology for Anomaly Detection and Incident Management in ICT Systems.** ESWC'24.

An ontology for Dynamic ICT systems



L. Tailhardat et al. **NORIA-O: An Ontology for Anomaly Detection and Incident Management in ICT Systems.** ESWC'24.

An ontology for Dynamic ICT systems

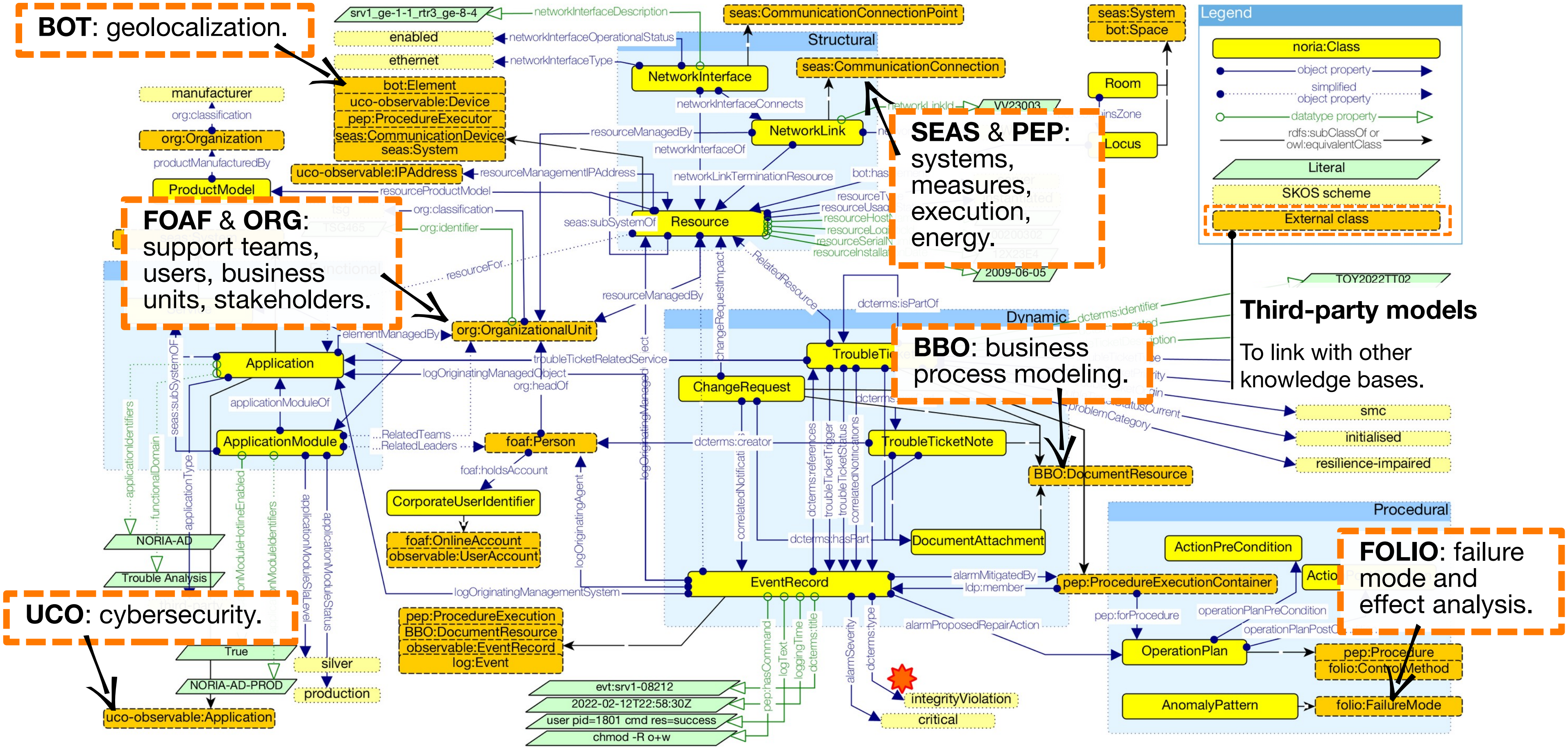


Controlled vocabulary

To handle heterogeneous data at the data value level, regardless of its origin.

L. Tailhardat et al. **NORIA-O: An Ontology for Anomaly Detection and Incident Management in ICT Systems.** ESWC'24.

An ontology for Dynamic ICT systems



L. Tailhardat et al. **NORIA-O: An Ontology for Anomaly Detection and Incident Management in ICT Systems.** ESWC'24.

Thanks !