

# Anomaly Detection using Knowledge Graphs and Synergistic Reasoning

EGC HYSNIC 2025



Imagined by you generated by Dino

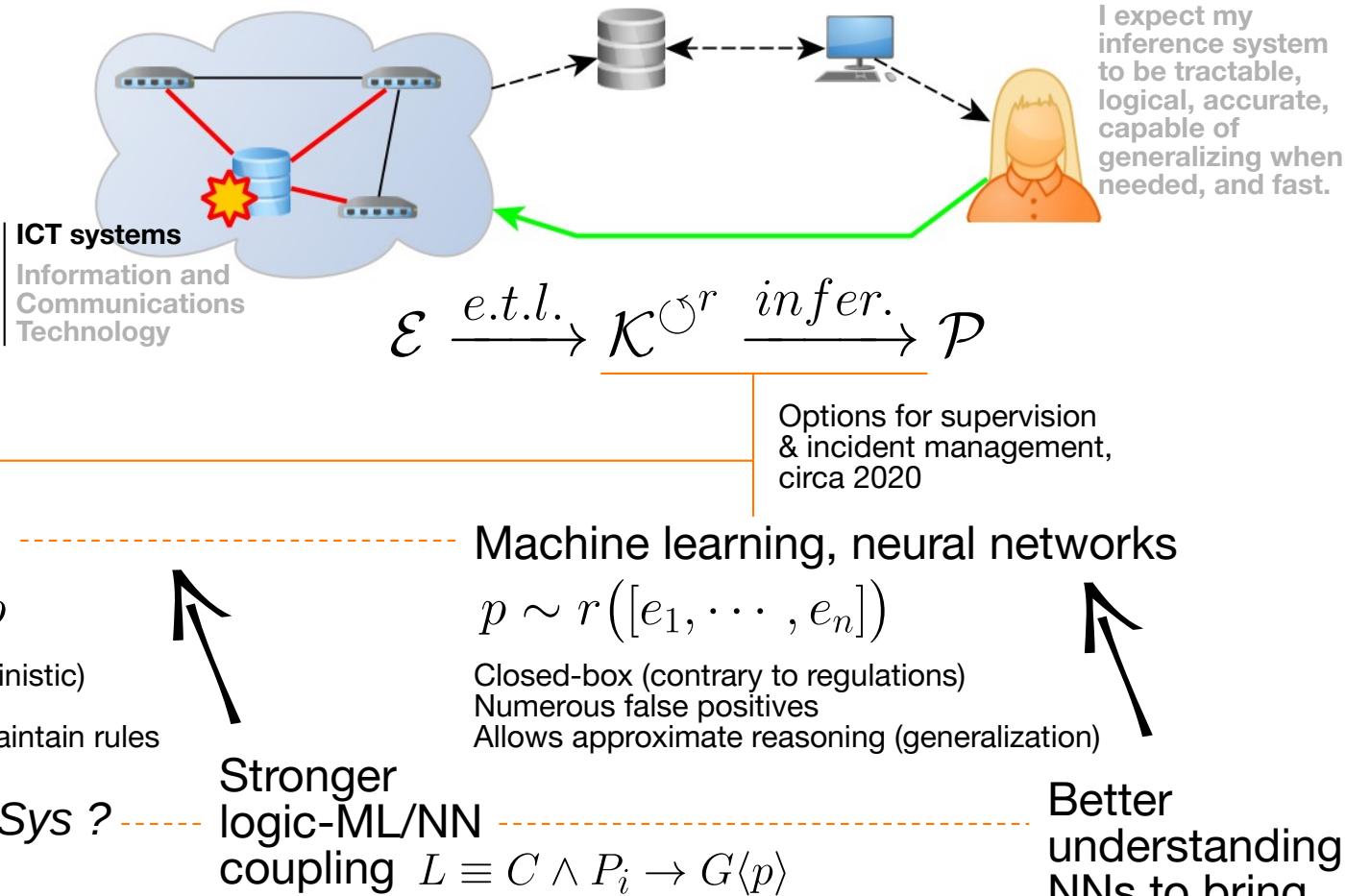
Lionel TAILHARDAT

[lionel.tailhardat@orange.com](mailto:lionel.tailhardat@orange.com)

<https://genears.github.io/>



# Back to the future



# | Building bridges, exploring possibilities

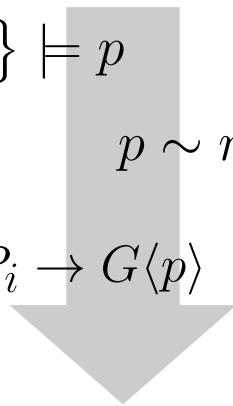
$$\{e_1, \dots, e_n\} \models p$$

$$p \sim r([e_1, \dots, e_n])$$

$$L \equiv C \wedge P_i \rightarrow G\langle p \rangle$$

Business rules and induction  
techniques... a possible synergy?

*Intermediary steps / findings are required!*



Anomaly Detection using  
Knowledge Graphs and  
Synergistic Reasoning  
Application to Network Management and  
Cyber Security

PhD Candidate - September 30, 2024  
Lionel TAILHARDAT

Oscar CORCHO, reviewer  
Oliver FESTOR, reviewer  
Anastasia DIMOU, examiner  
Adlen KSENTINI, examiner

Ulrich FINGER, thesis director  
Raphaël TRONCY, thesis co-director  
Yoan CHABOT, thesis co-director



<https://theses.fr/2024SORUS293>

# Networks and us

Today, Alice's FluffyChat messenger cannot reach Bob's and Charlie's ...

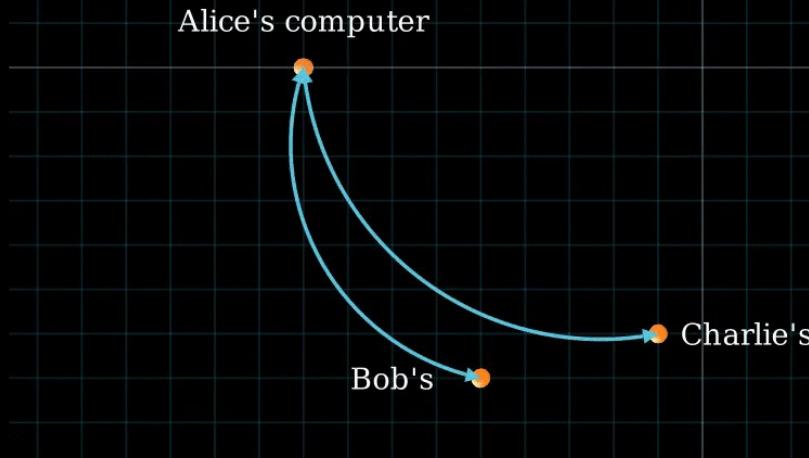
... who's to blame ?

- Wrong action
- Bug in the Matrix protocol
- Spontaneous network fault
- Cyberattack

Let's ask Susie, a network & security supervision expert ...

# Networks and us

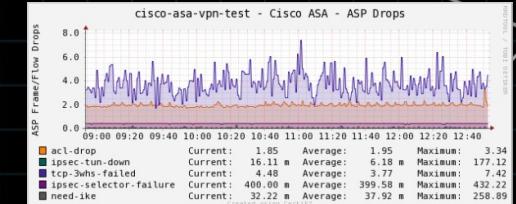
The network is more complex than we may think, from both a **structural**, **functional**, and **dynamic** perspectives ...



... we must have a bird's eye view for situation understanding, and selecting the appropriate **procedure** to solve the issue.

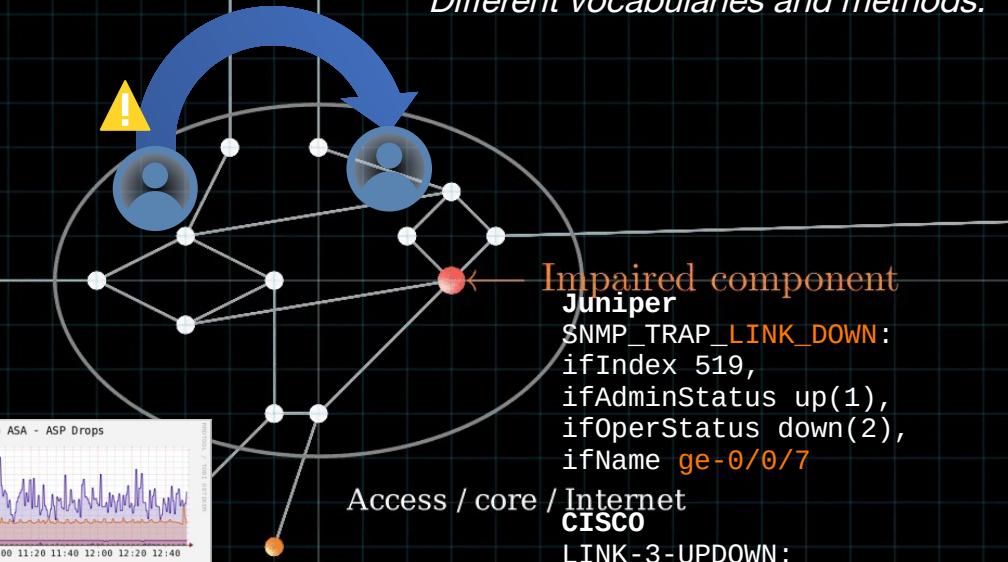
# Networks and us

A single bird cannot grasp everything due to the coexistence and interplay of multiple ...



*Trend analysis and change point detection in a time series.*

Organizations and operator profiles  
*Observability issues.  
Different vocabularies and methods.*

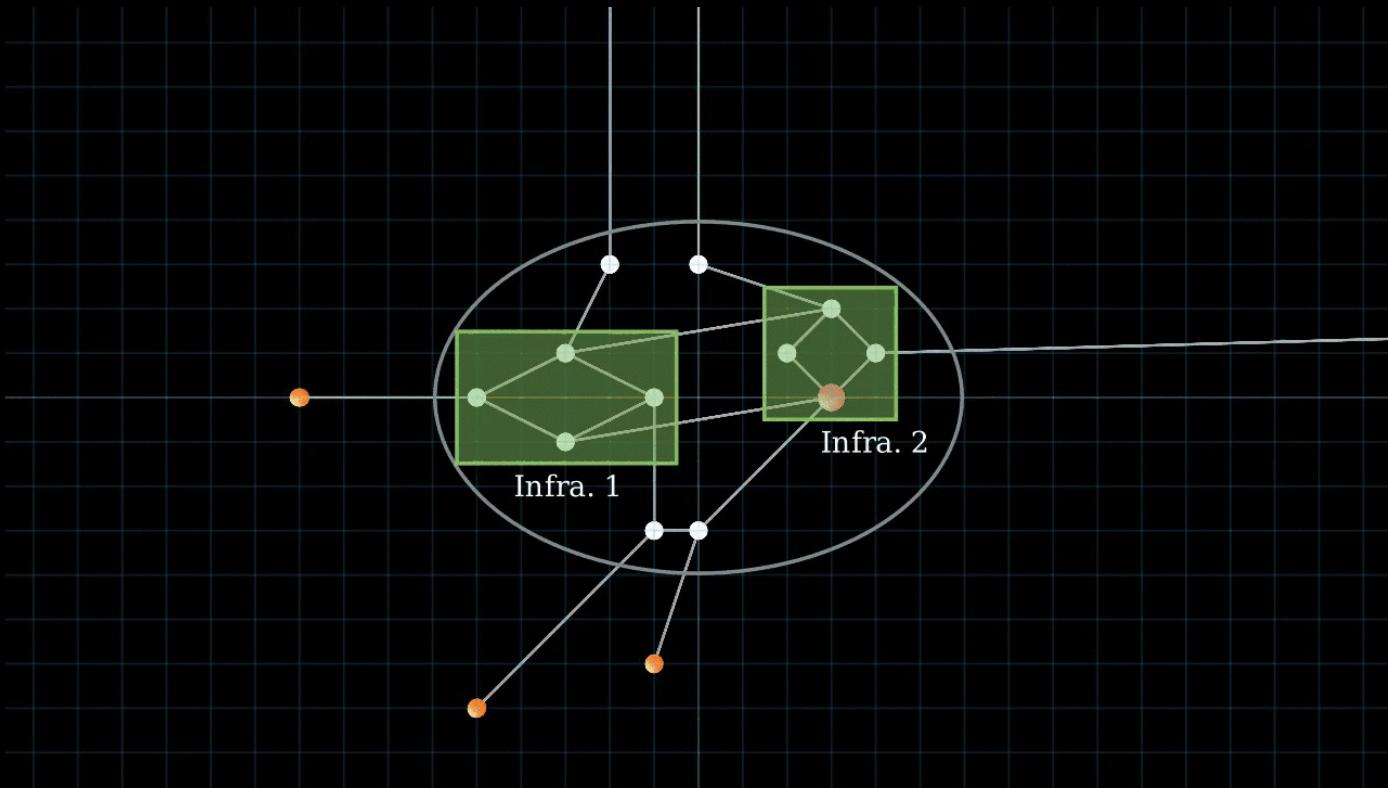


Access / core / Internet  
**CISCO**  
*Rule-based state change detection in parsed logs.*

Technologies, device manufacturers, configurations, and monitoring systems  
*Heterogeneity in knowledge representations and semantics of phenomena.  
Limited decision support code reuse and inference aggregation.*

# Networks and us

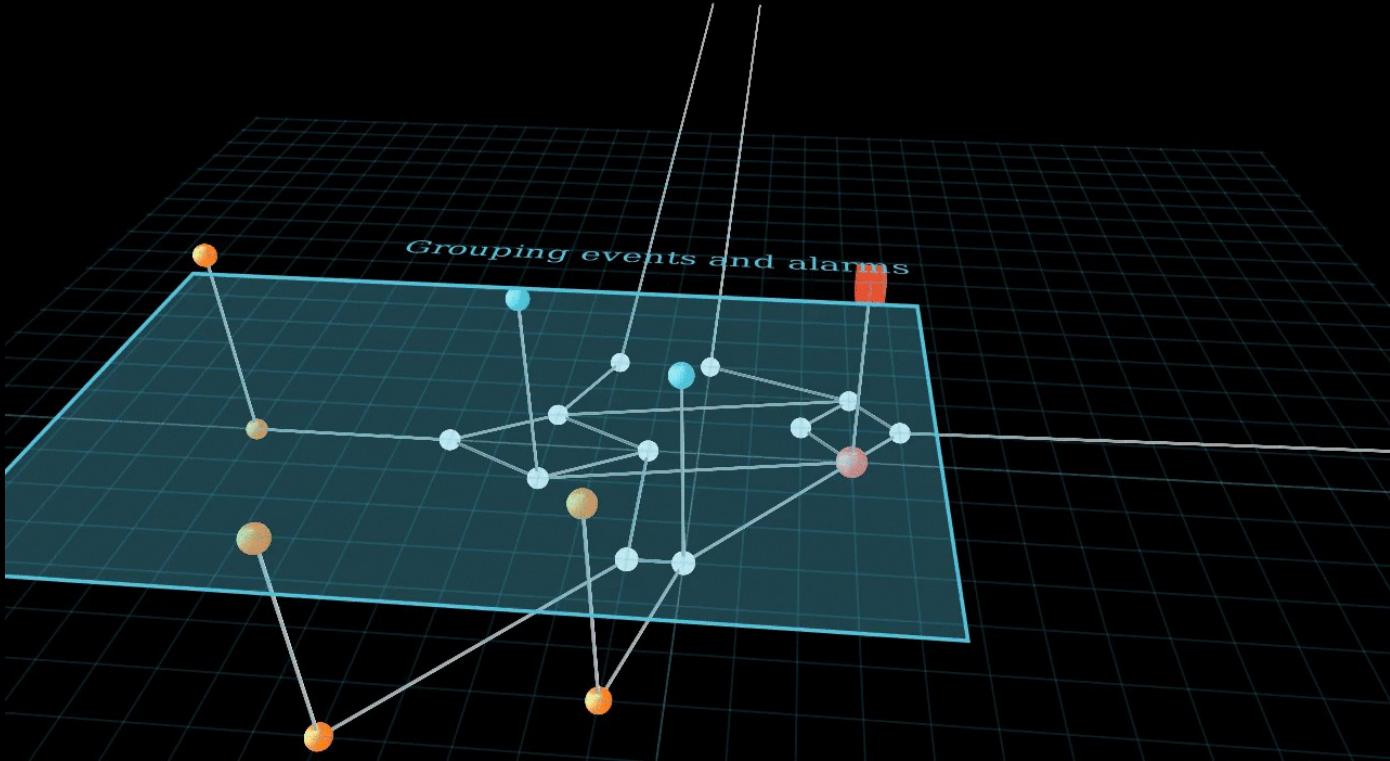
Could therefore be interesting to have a unified view of the assets by **handling heterogeneous data...**



... and also of their global **behavior!**

# Networks and us

... which could help us fully capture an **reason about an incident context**, including its **internal logic**.



Anomal Detection (AD) and Root Cause Analysis (RCA) of complex situations  
*Increase in operational efficiency.  
Lower cognitive effort.*

→ Improving the design of ICT systems  
*Knowledge capitalization on the systems behaviors.  
Knowledge sharing across operators and designers.*

# 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?

## Constraints on the internal representation of data and knowledge

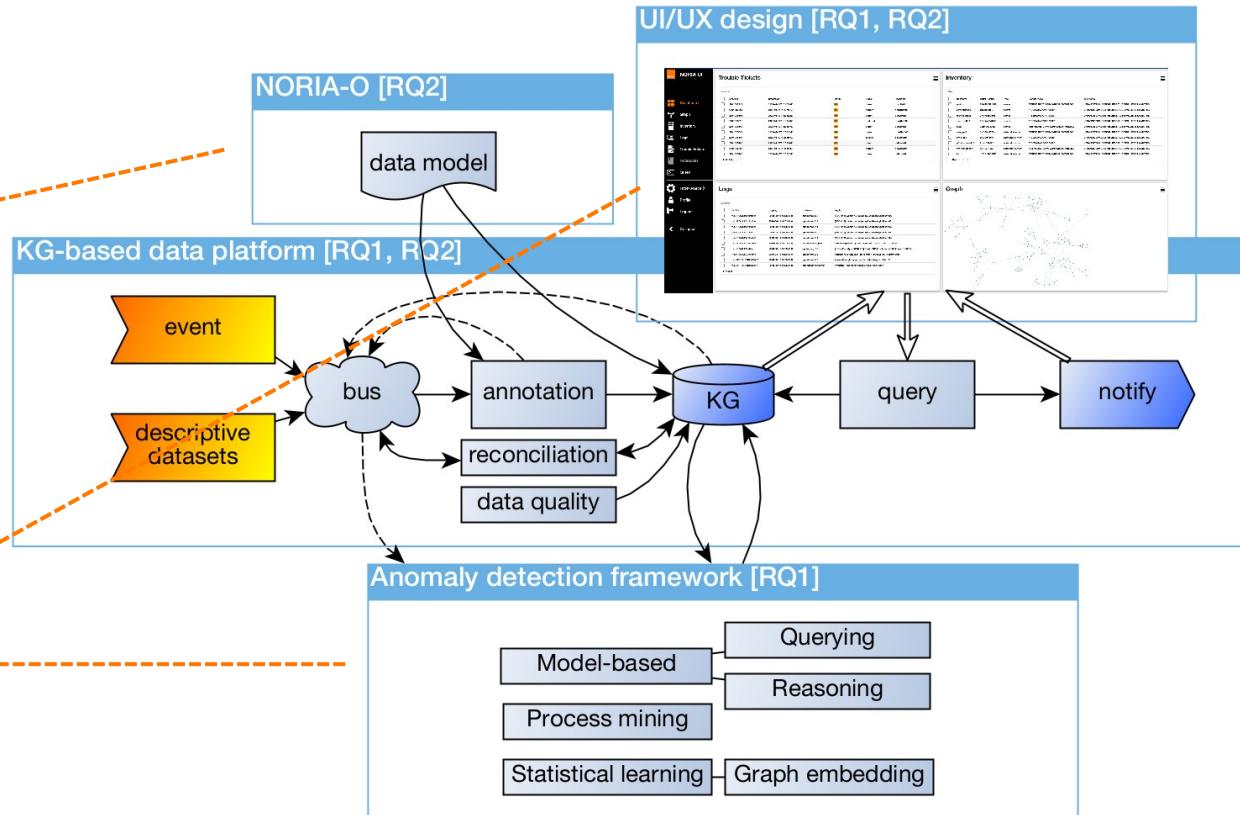
Can human operators and decision support AI agents use the same Knowledge Representation (KR) of ICT systems for anomaly detection and knowledge management, that KR being subject to computation efficiency and interpretability?

RQ. 2

# Research Roadmap

Part I

Building a **graph** for dynamic ICT systems



RQ. 1 - Anomaly model production & utilization with heterogeneous data  
RQ. 2 - Constraints on the internal representation of data and knowledge

# Building a graph for dynamic ICT systems

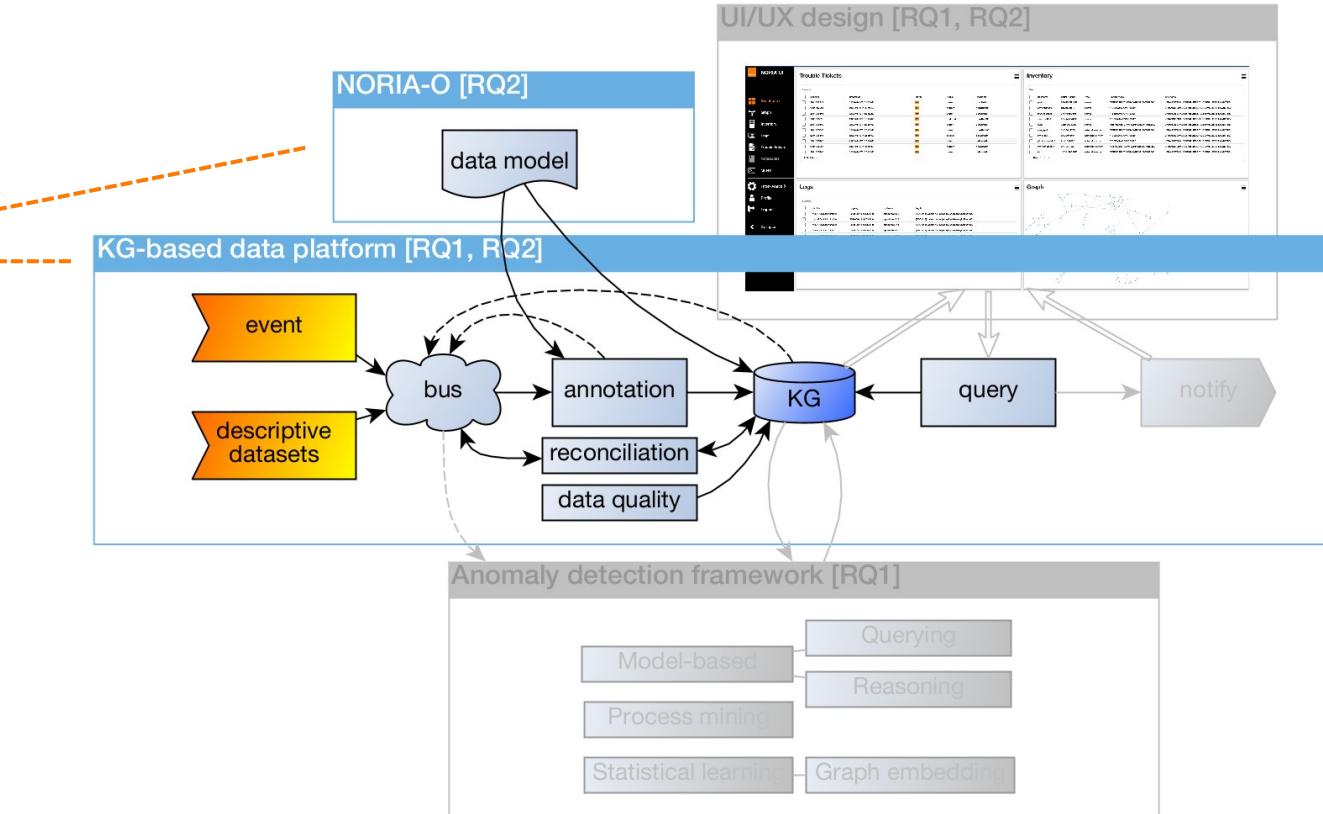
Part I



# Research Roadmap

Part I

## Building a graph for dynamic ICT systems



RQ. 1 - Anomaly model production & utilization with heterogeneous data  
RQ. 2 - Constraints on the internal representation of data and knowledge

# Analysis of Semantic Models

95 references analyzed: to what extent the set of models for each application domain theoretically aligns with the targeted discourse domain ?

Theme	MC	St. %	Fu. %	Dy. %	Pr. %	F0 %	F1 %	F2 %	F3 %	F4 %
Generic	18	0,0	11,1	55,6	38,9	<b>33,3</b>	33,3	27,8	5,6	0,0
CyberSec	11	54,5	54,5	63,6	81,8	0,0	36,4	18,2	0,0	<b>45,5</b>
SE-SI	9	<b>88,9</b>	<b>66,7</b>	55,6	44,4	0,0	11,1	<b>44,4</b>	22,2	22,2
Net-IT	7	71,4	42,9	28,6	28,6	0,0	<b>42,9</b>	42,9	14,3	0,0
Process modeling	4	50,0	25,0	<b>75,0</b>	<b>100,0</b>	0,0	25,0	25,0	<b>25,0</b>	25,0
Health Science	1	<i>100,0</i>	0,0	0,0	<i>100,0</i>	0,0	0,0	<i>100,0</i>	0,0	0,0
Overall	50	44,0	36,0	<b>54,0</b>	<b>54,0</b>	12,0	30,0	<b>32,0</b>	10,0	16,0

MC: model count ; St.: structural, Fu.: functional, Dy.: dynamic, Pr.: procedural

St.%, Fu.%, Dy.%, Pr.%: proportion of models for which the facet has been identified

Fx%: expressiveness of the models by comparing the proportion of models that meet 0, 1, 2, 3, or 4 facets.

 Vandebussche et al. **Linked Open Vocabularies (LOV): A Gateway to Reusable Semantic Vocabularies on the Web**. SWJ, 2017.

 Rivadeneira et al. **Cybersecurity Ontologies: A Systematic Literature Review**. ReCIBE, 2020.

 Abu-Salih. **Domain-specific knowledge graphs: A survey**. Journal of Network and Computer Applications, 2021.

# Analysis of Semantic Models

Six primary application domains (theme), with varying proportions of available models and model characteristics...

→ Q: to what extent the set of models for each application domain aligns with the targeted discourse domain ?

Theme	MC	St. %	Fu. %	Dy. %	Pr. %	F0 %	F1 %	F2 %	F3 %	F4 %
Generic	18	0,0	11,1	55,6	38,9	<b>33,3</b>	33,3	27,8	5,6	0,0
CyberSec	11	54,5	54,5	63,6	81,8	0,0	36,4	18,2	0,0	<b>45,5</b>
SE-SI	9	<b>88,9</b>	<b>66,7</b>	55,6	44,4	0,0	11,1	<b>44,4</b>	22,2	22,2
Net-IT	7	71,4	42,9	28,6	28,6	0,0	<b>42,9</b>	42,9	14,3	0,0
Process modeling	4	50,0	25,0	<b>75,0</b>	<b>100,0</b>	0,0	25,0	25,0	<b>25,0</b>	25,0
Health Science	1	<b>100,0</b>	0,0	0,0	<b>100,0</b>	0,0	0,0	<b>100,0</b>	0,0	0,0
Overall	50	44,0	36,0	<b>54,0</b>	<b>54,0</b>	12,0	30,0	<b>32,0</b>	10,0	16,0

MC: model count ; St.: structural; Fu.: functional; Dy.: dynamic, Pr.: procedural

50/95 with implementation based on Semantic Web technologies.

The 45 others did not have an implementation.

# Analysis of Semantic Models

95 references analyzed: to what extent the domain theoretically aligns with the target discourse domain for each application

Facet coverage varies across the different groups of models.

Low coupling between facets.

Theme	MC	St. %	Fu. %	Dy. %	Pr. %	F0 %	F1 %	F2 %	F3 %	F4 %
Generic	18	0,0	11,1	55,6	38,9	<b>33,3</b>	33,3	27,8	5,6	0,0
CyberSec	11	54,5	54,5	63,6	81,8	0,0	36,4	18,2	0,0	<b>45,5</b>
SE-SI	9	<b>88,9</b>	<b>66,7</b>	55,6	44,4	0,0	11,1	<b>44,4</b>	22,2	22,2
Net-IT	7	71,4	42,9	28,6	28,6	0,0	<b>42,9</b>	42,9	14,3	0,0
Process modeling	4	50,0	25,0	<b>75,0</b>	<b>100,0</b>	0,0	25,0	25,0	<b>25,0</b>	25,0
Health Science	1	<b>100,0</b>	0,0	0,0	<b>100,0</b>	0,0	0,0	<b>100,0</b>	0,0	0,0
Overall	50	44,0	36,0	<b>54,0</b>	<b>54,0</b>	12,0	30,0	<b>32,0</b>	10,0	16,0

MC: model count ; St.: structural, Fu.: functional, Dy.: dynamic, Pr.: procedural

St.%, Fu.%, Dy.%, Pr.%: proportion of models for which the facet has been identified

Fx%: expressiveness of the models by comparing the proportion of models that meet 0, 1, 2, 3, or 4 facets.

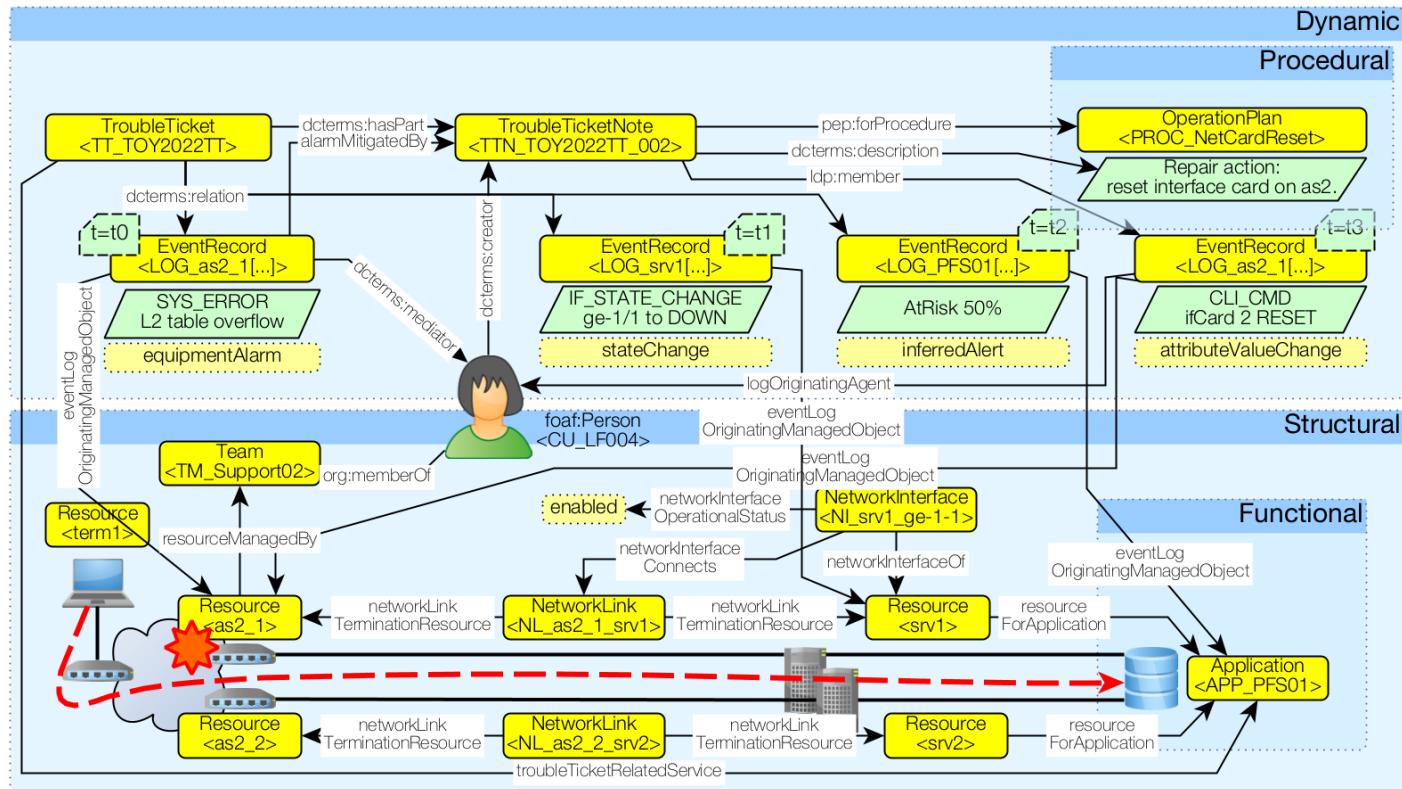
## Challenges in Knowledge Representation & Reasoning (KRR)

Potential difficulties in precisely allowing for reasoning on the **interplay** between **network architecture** and its **operation**.



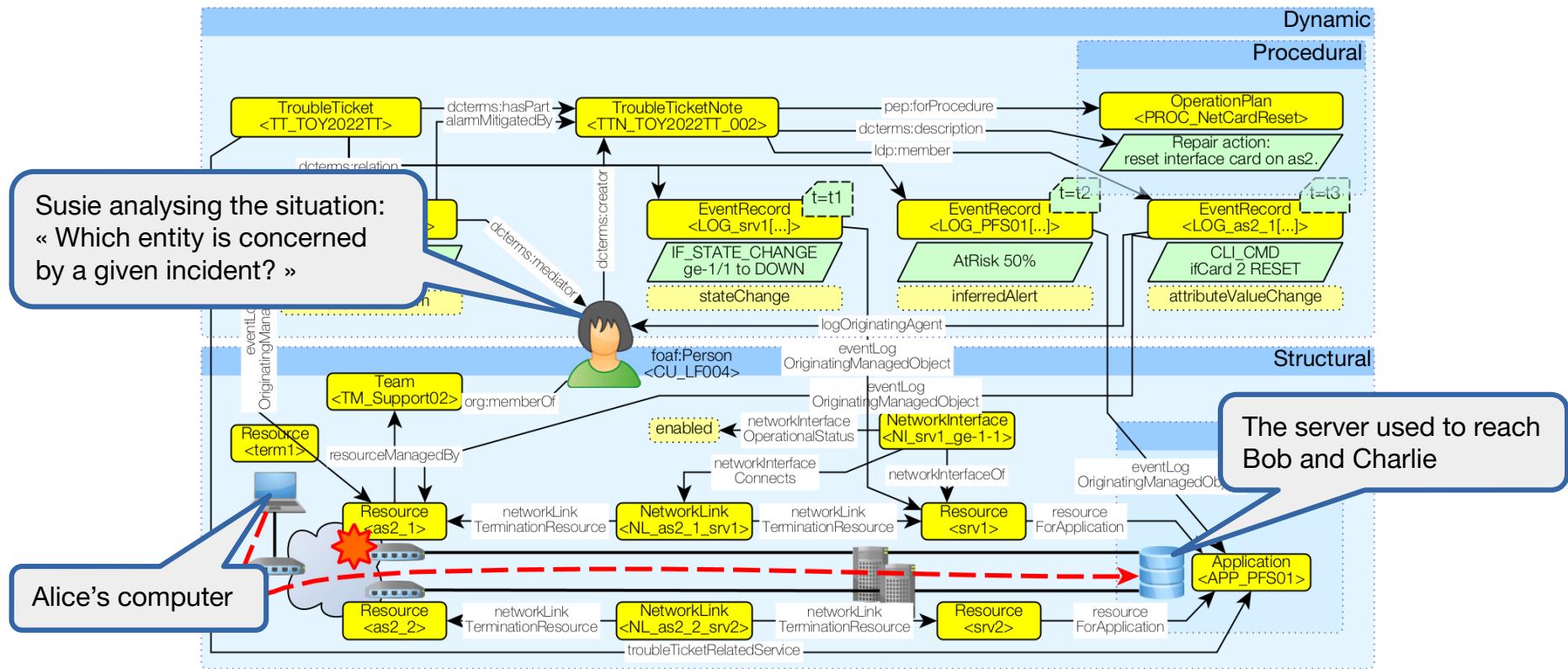
# Knowledge Graphs ?

Enable data analysis and inference techniques to reason about the **context** of represented objects while handling **heterogeneous data**.

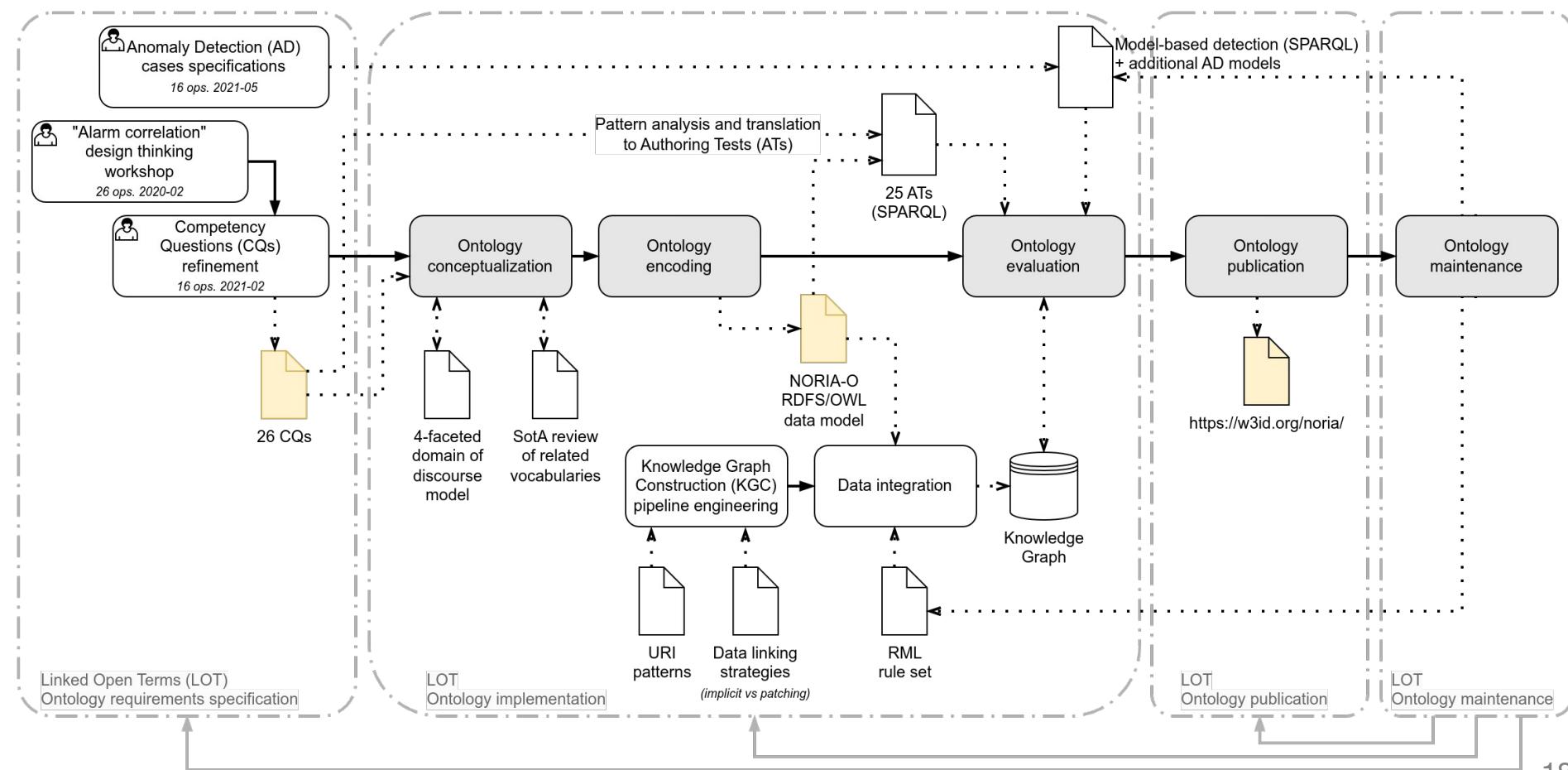


# Knowledge Graphs ?

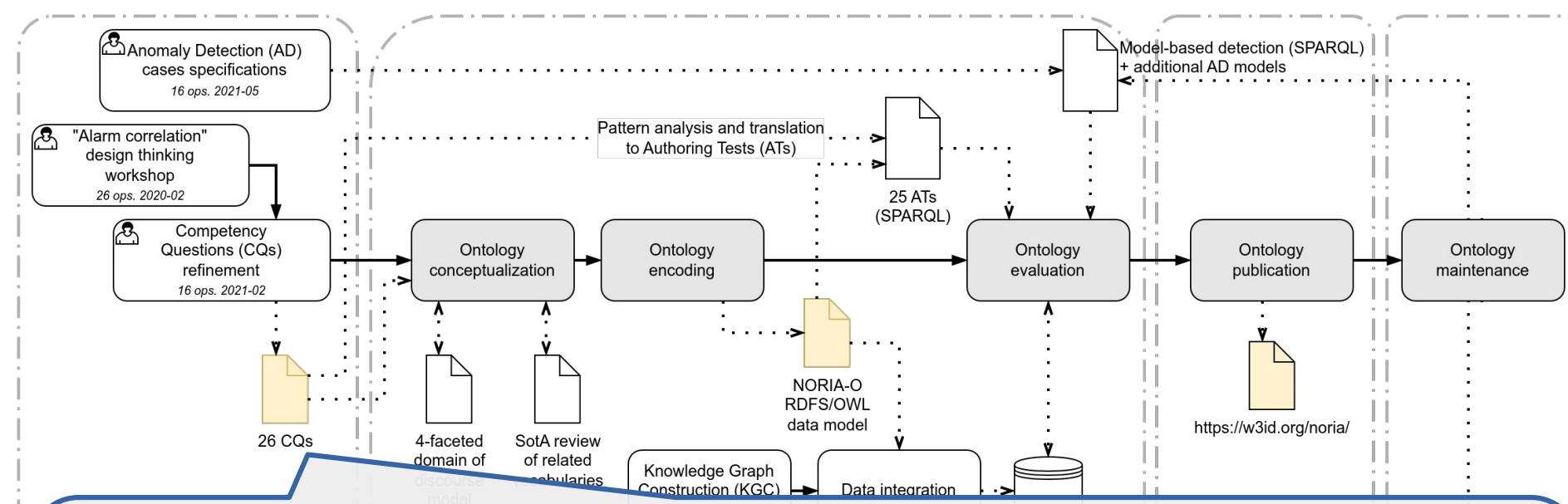
Enable data analysis and inference techniques to reason about the context of represented objects while handling heterogeneous data.



# Knowledge Engineering



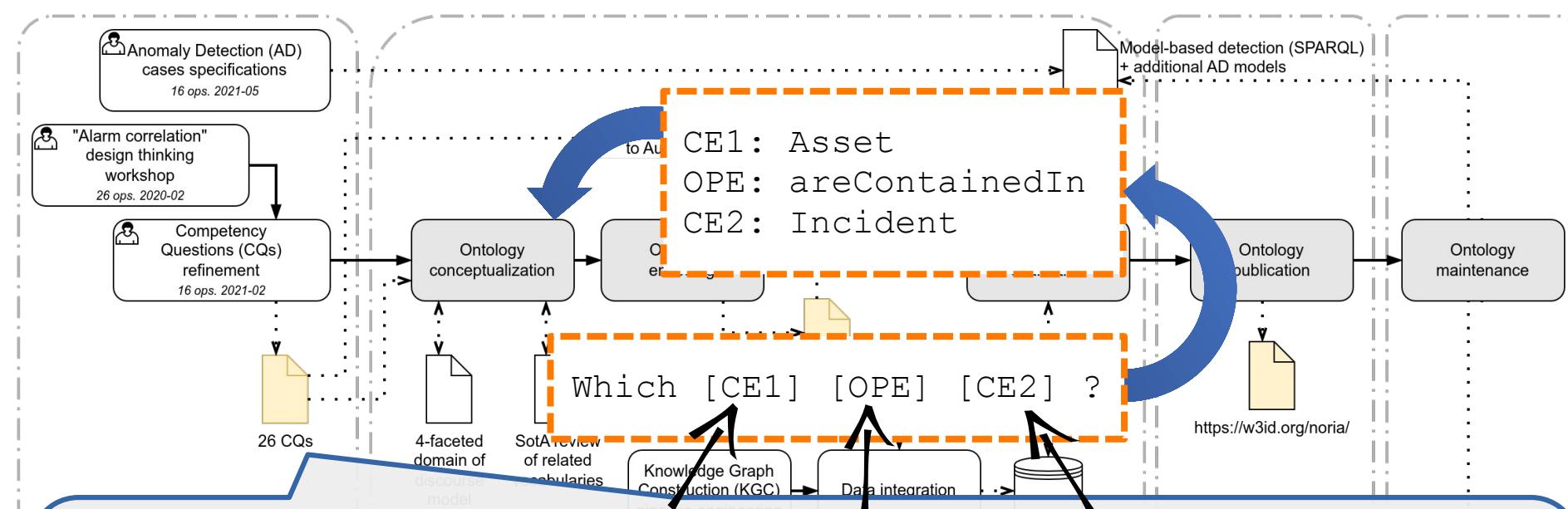
# Knowledge Engineering



- « Which entity (resource/application/site) is concerned by a given incident? » (CQ1)
- « What was the root cause of the incident? » (CQ11)
- « What is the financial cost of this incident if it occurs? » (CQ23)
- « What are the vulnerabilities and the associated risk levels of this infrastructure? » (CQ25)



# Knowledge Engineering



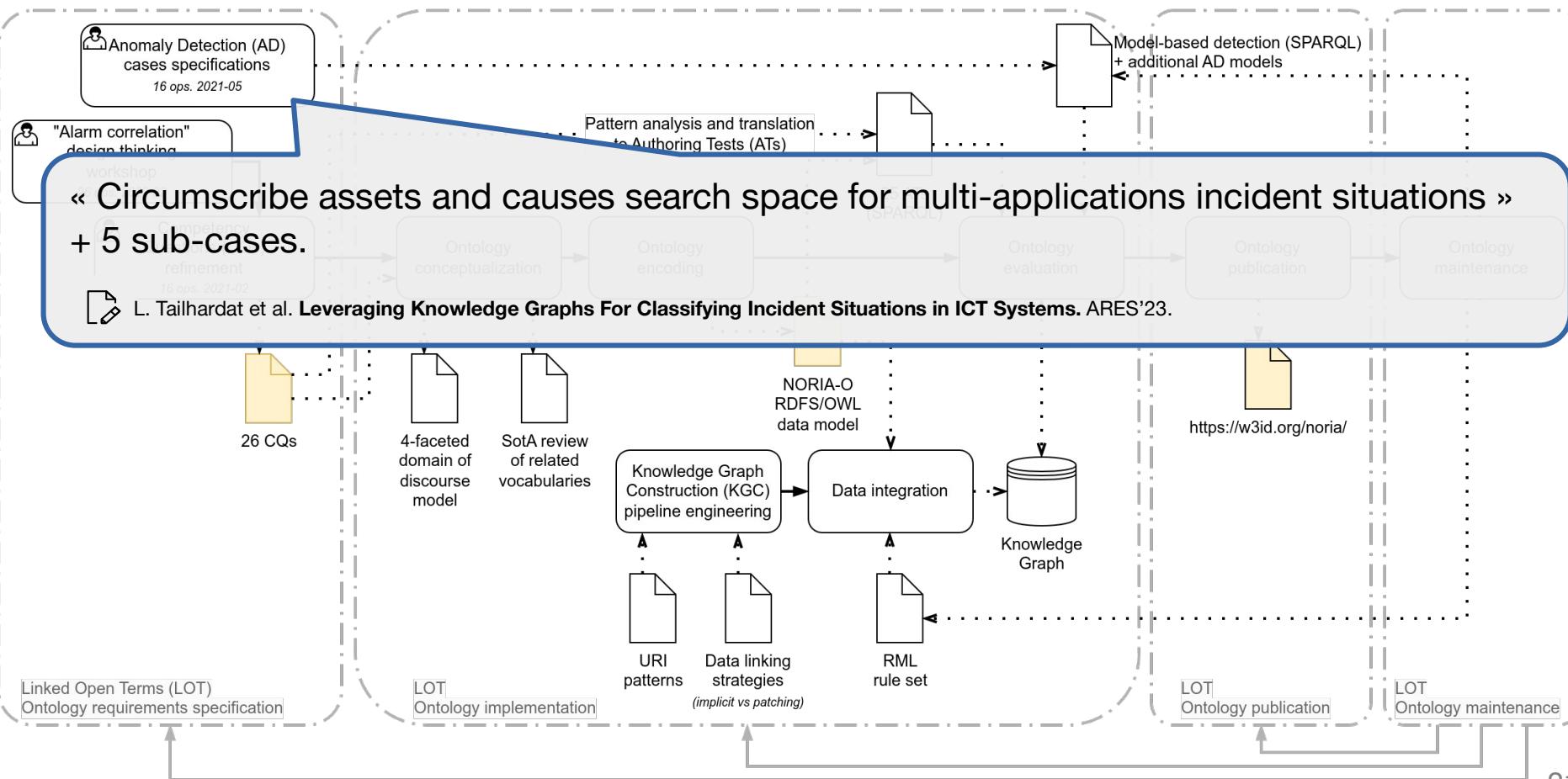
« Which [entity (resource/application/site)] is [concerned] by a given [incident]? » (CQ1)

« What was the root cause of the incident? » (CQ11)

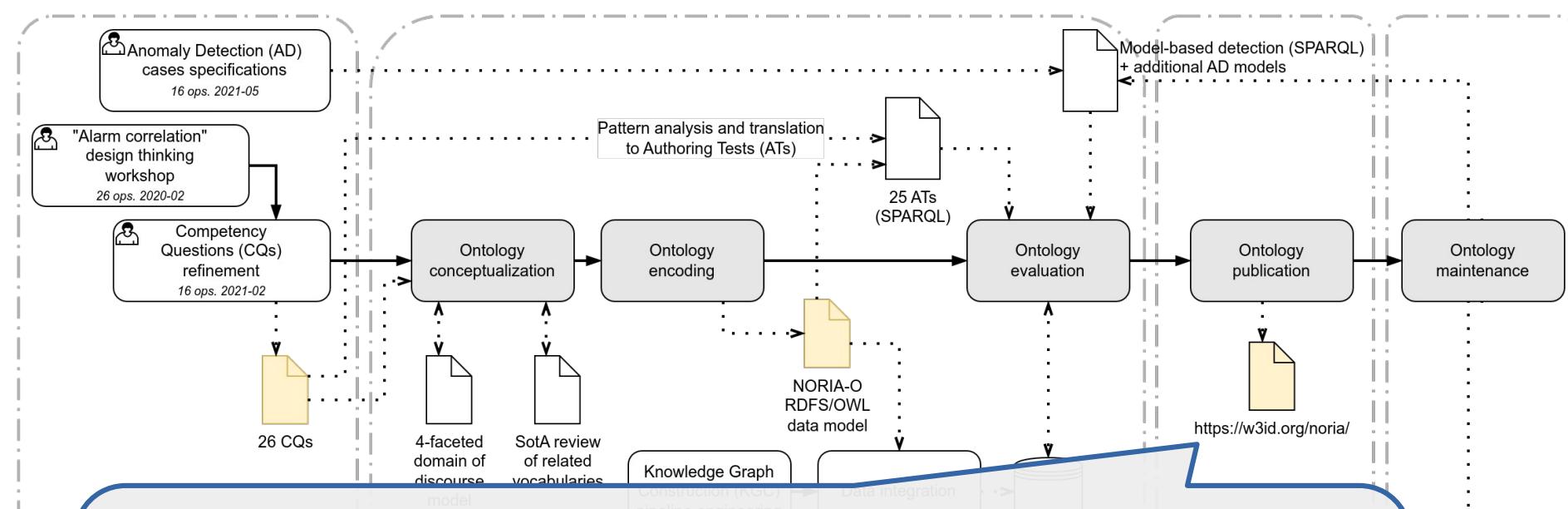
« What is the financial cost of this incident if it occurs? » (CQ23)

« What are the vulnerabilities and the associated risk levels of this infrastructure? » (CQ25)

# Knowledge Engineering



# Knowledge Engineering

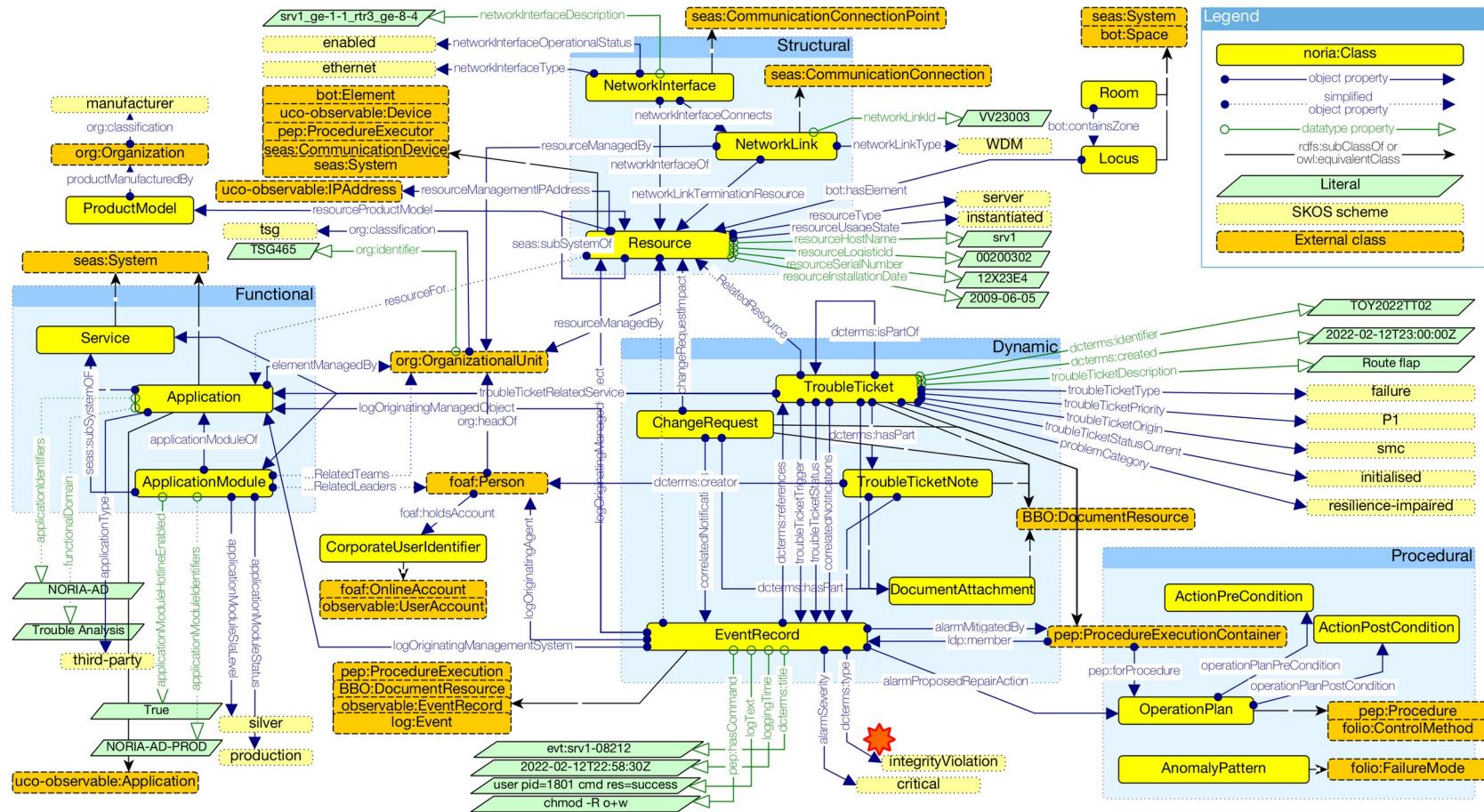


NORIA-O v0.3 - open source release under BSD-4 license

- Implementation: RDFS/OWL-2 + SKOS (controlled vocabulary).
- Statistics: 59 classes, 107 object properties, 71 datatype properties, 57 SKOS ConceptSchemes, 264 SKOS Concepts.
- Four facets: Structural, Functional, Dynamics, Procedural.



# An ontology for Dynamic ICT systems



# An ontology for Dynamic ICT systems

Alice's computer, the server used to reach Bob and Charlie, etc.

The instant messaging service for Alice to reach out to Bob and Charlie.

A document to follow-up on the incident « Alice's computer cannot reach Bob's and Charlie's »

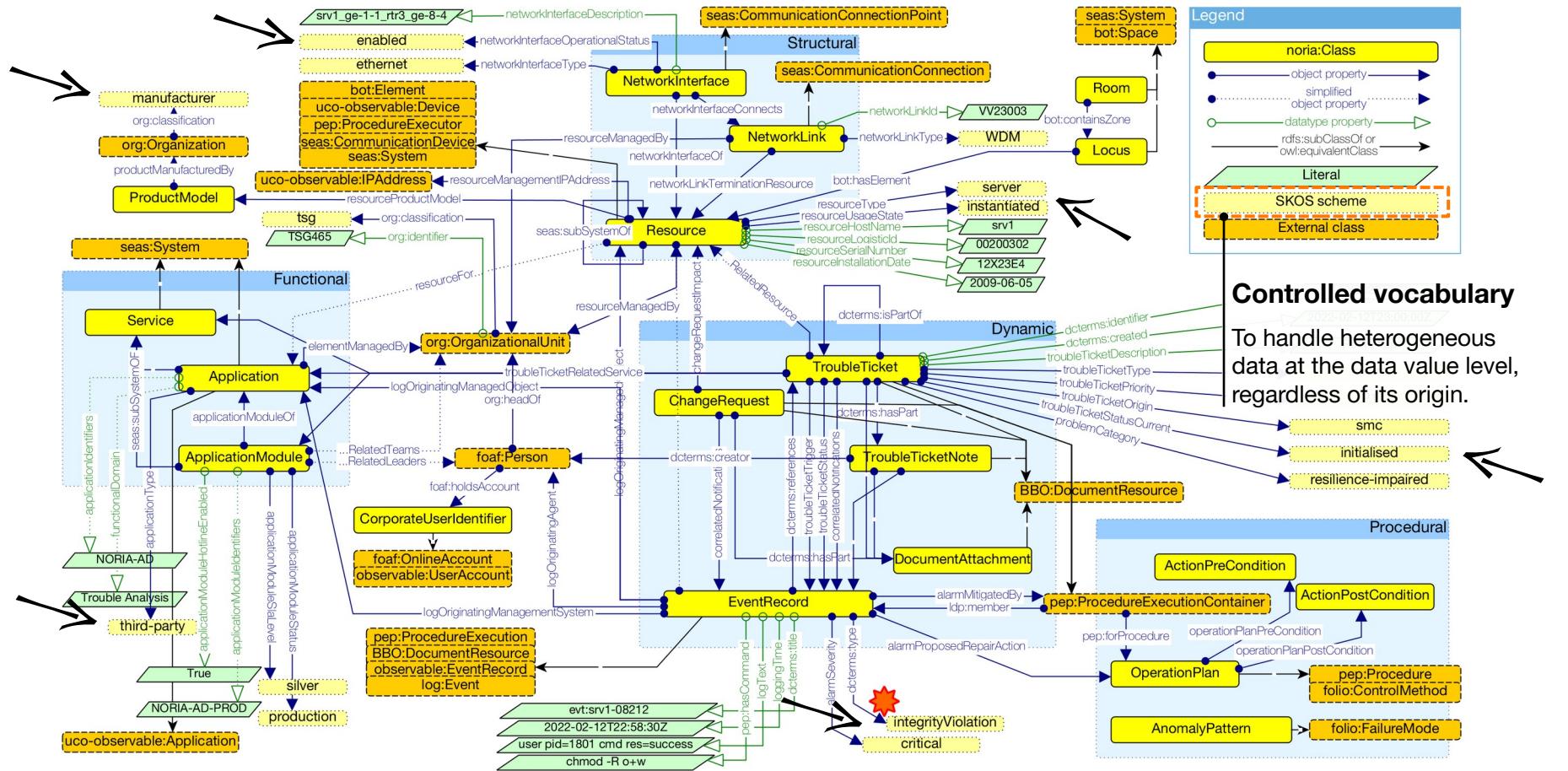
Susie and other network stakeholders.

Alarms and logs from the network that reflect the impairment of the instant messaging service.

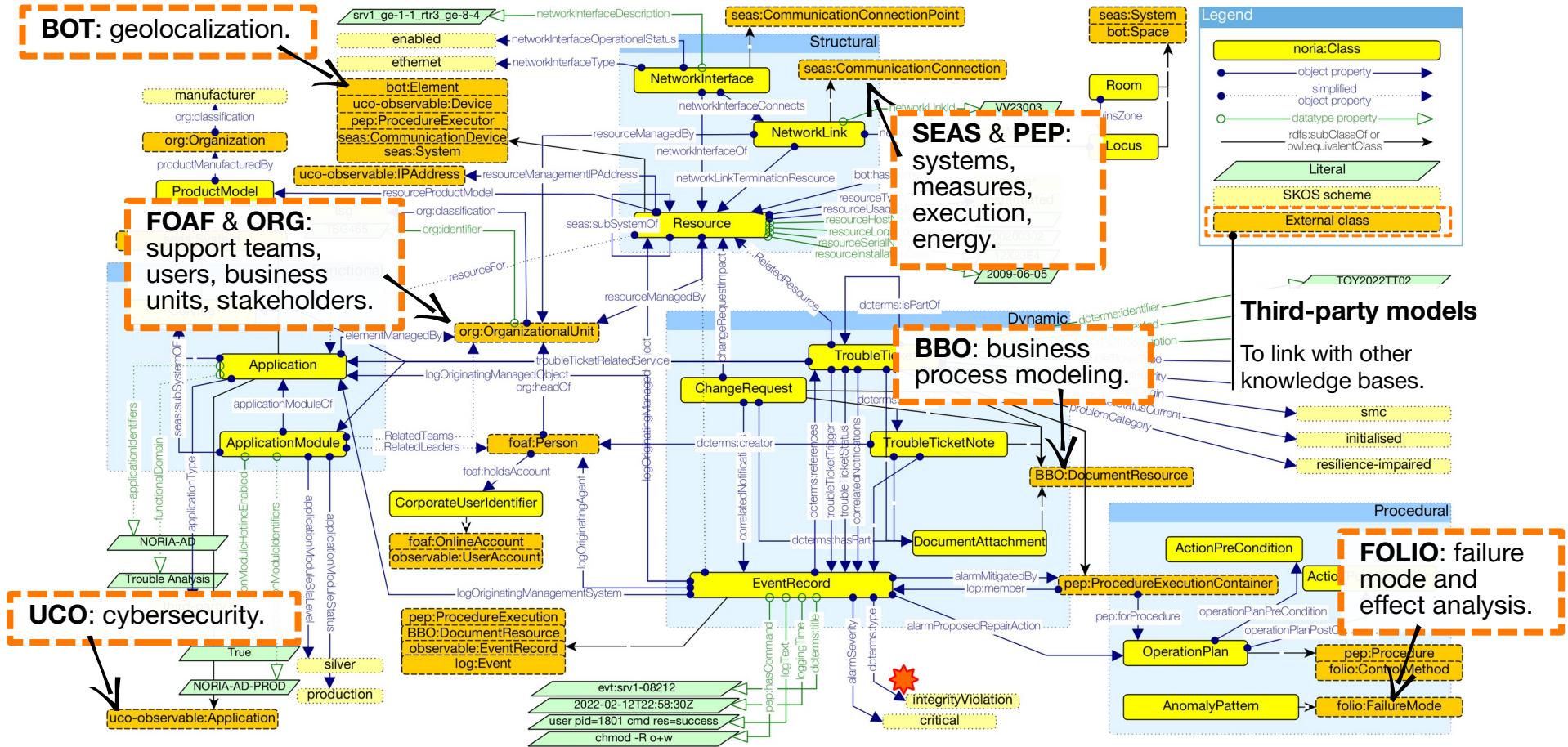
Expert knowledge for root cause analysis (RCA) and incident response.



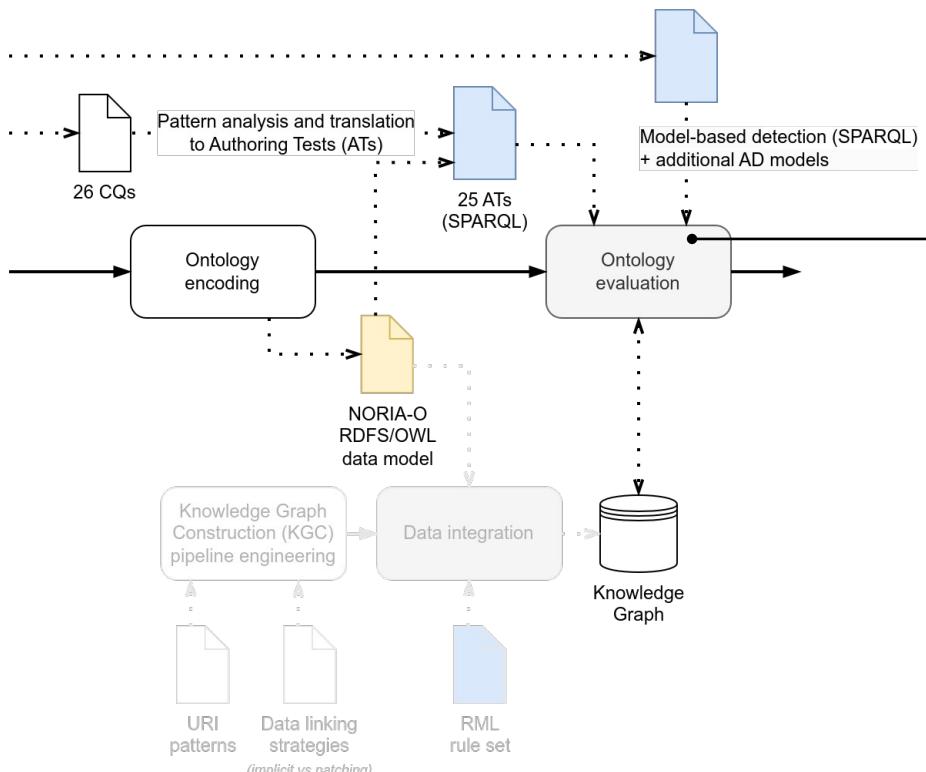
# An ontology for Dynamic ICT systems



# An ontology for Dynamic ICT systems



# Evaluation and Results

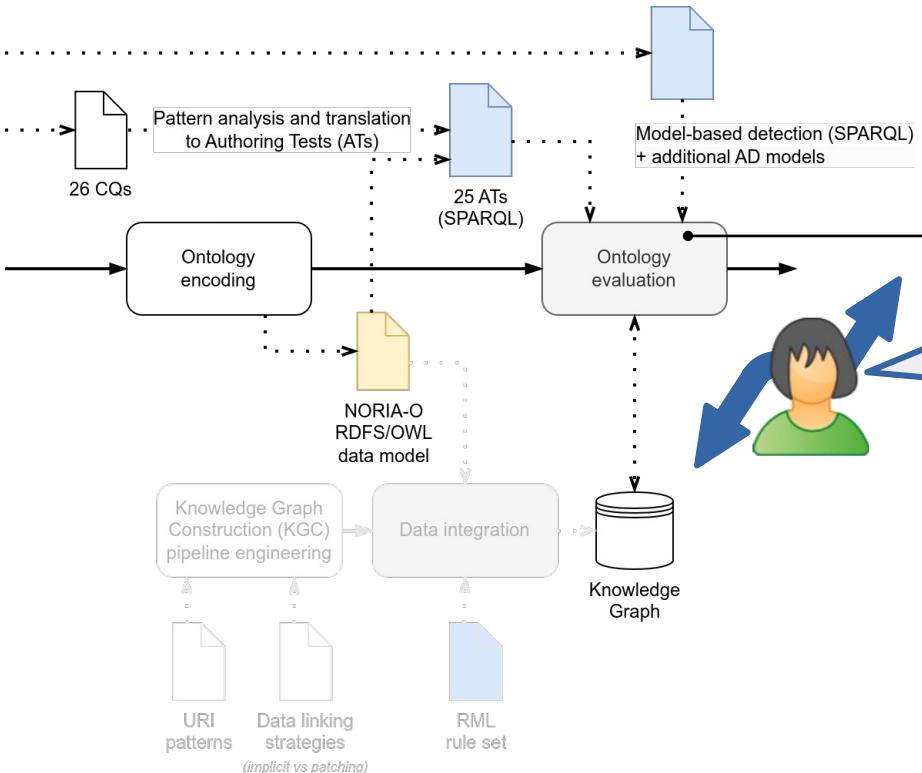


## Authoring Tests for NORIA-O [RQ. 2]

- ✓ 16/26 « OK » answered using a single or several simple SPARQL queries and the ontology.  
*“Which entity is concerned by a given incident?” (CQ1)*
- ✓ 9/26 « AI » require the implementation of more complex AI-based algorithms such as anomaly detection algorithms.  
*“What was the root cause of the incident?” (CQ11) → the explicit representation of alarms and logs associated with a given incident is not enough and needs to be enhanced with root cause analysis algorithms.*
- ✓ 1/26 « Extension » require the introduction of new concepts or relations via an extension of the NORIA-O model.  
*“What is the financial cost of this incident if it occurs?” (CQ23) → involves information about the cost of an incident.*

RQ. 1 - Anomaly model production & utilization with heterogeneous data  
RQ. 2 - Constraints on the internal representation of data and knowledge

# Evaluation and Results



## Authoring Tests for NORIA-O [RQ. 2]

- ✓ 16/26 « OK » answered using a single or several simple SPARQL queries and the ontology.

*“Which entity is concerned by a given incident?” (CQ1)*  
Ontologies bring **unified view of heterogeneous systems**, including their dynamics, in line with the way experts refer to their network.  
*“What is the explicit representation of alarms and logs associated with a given incident?” (CQ11) → the explicit representation of alarms and logs associated with a given incident is not enough and needs to be enhanced with root cause analysis algorithms.*

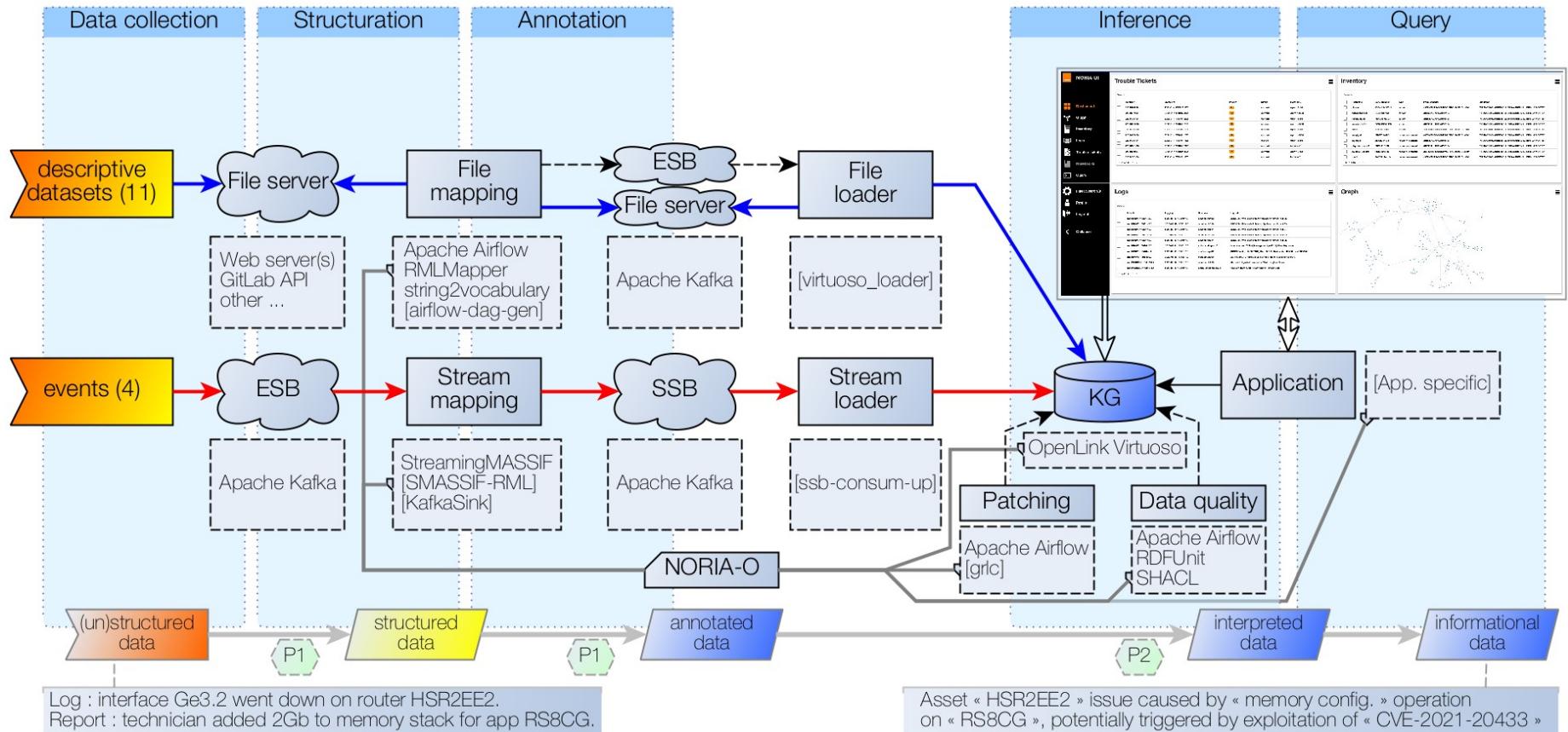
*“What are the vulnerabilities and the associated risk levels of this infrastructure?” (CQ25) → can be answered only by looking for non-desirable network topology shapes or relations to third-party cybersecurity vulnerability entities based on structure and security scanners.*

- ✓ 1/26 « Extension » require the introduction of new concepts or relations via an extension of the NORIA-O model.

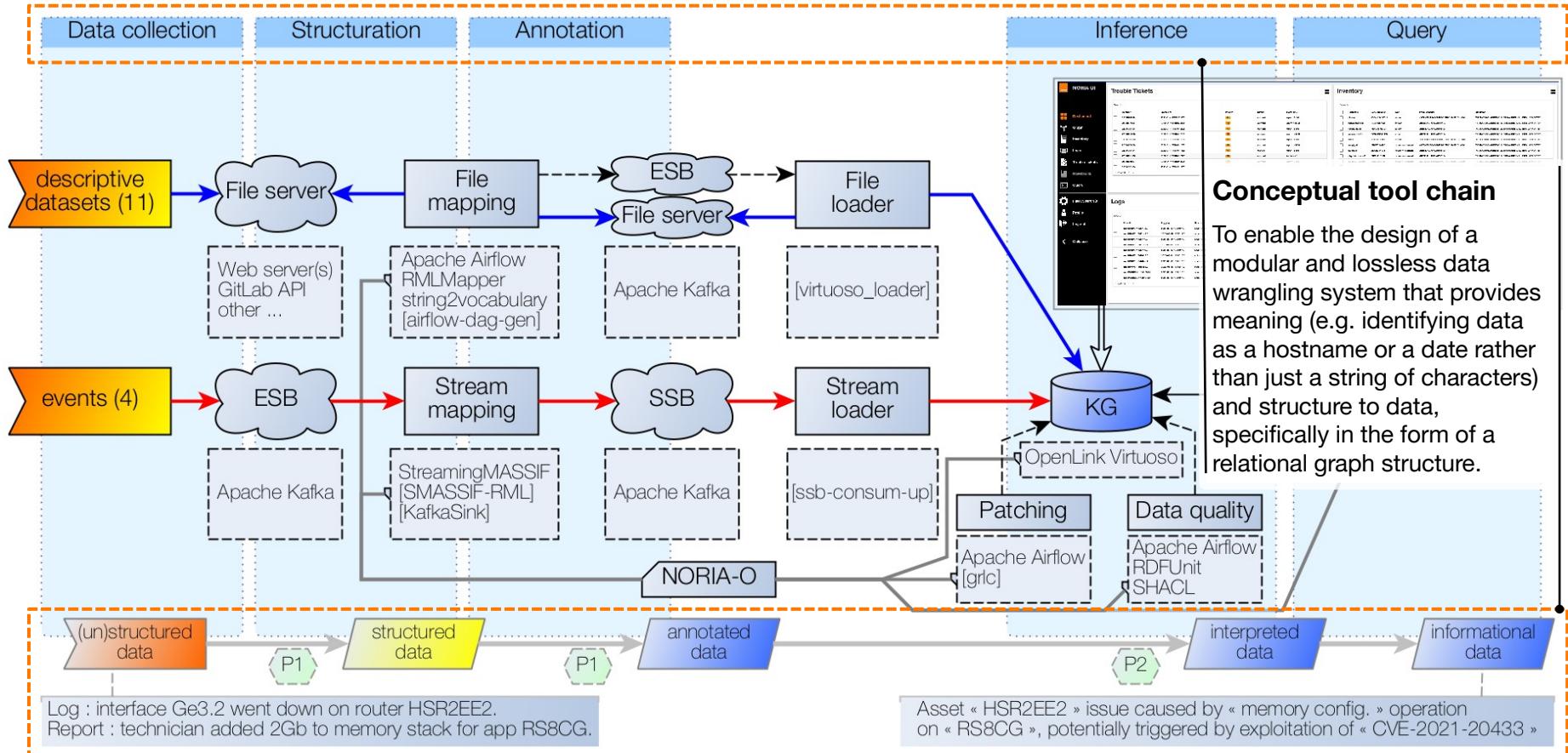
*“What is the financial cost of this incident if it occurs?” (CQ23) → involves information about the cost of an incident.*

RQ. 1 - Anomaly model production & utilization with heterogeneous data  
RQ. 2 - Constraints on the internal representation of data and knowledge

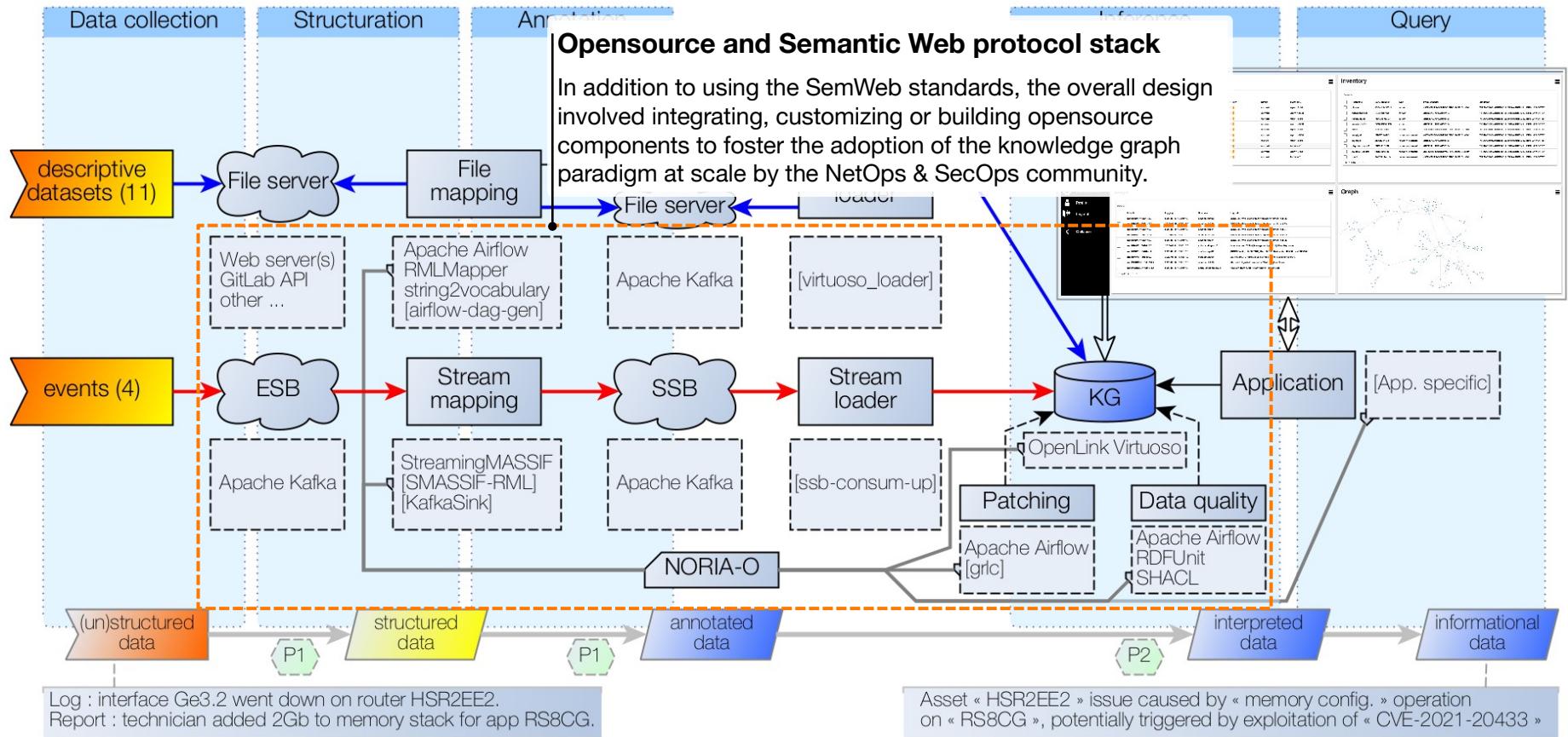
# Knowledge Graph Construction



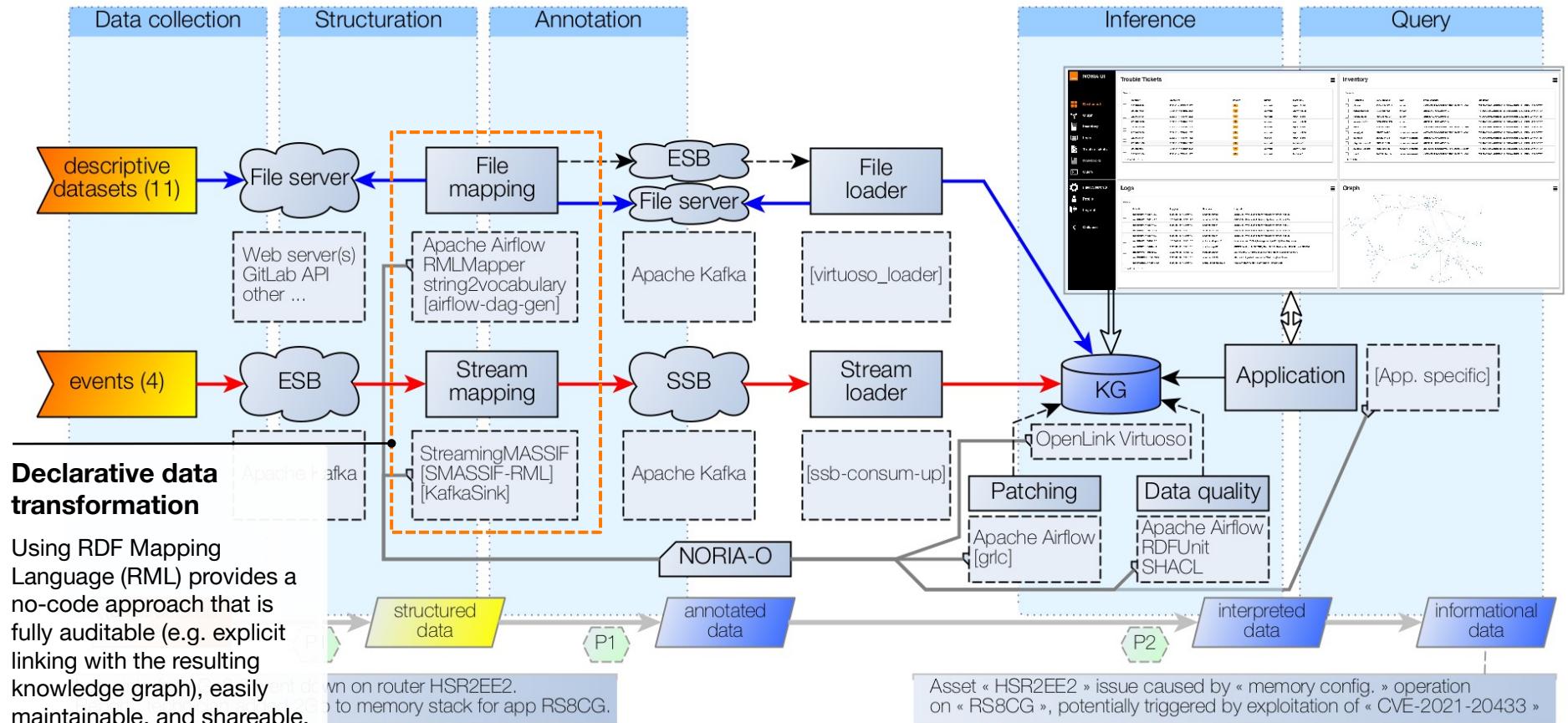
# Knowledge Graph Construction



# Knowledge Graph Construction

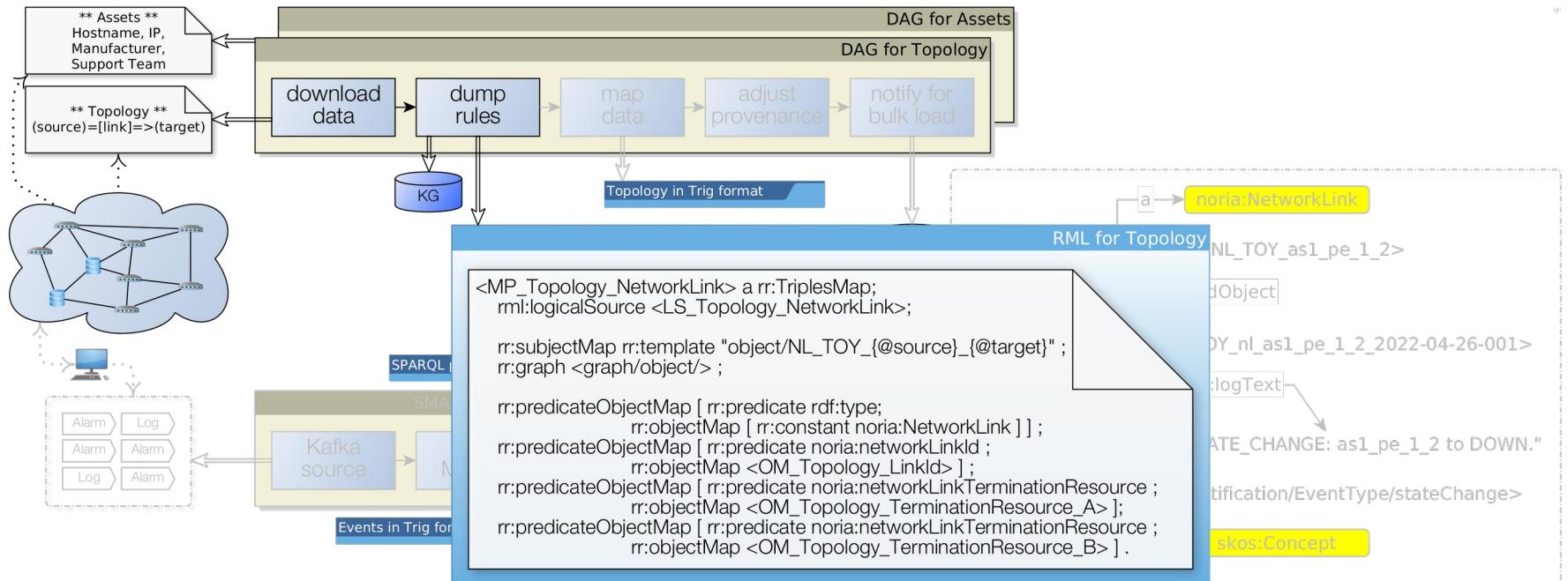


# Knowledge Graph Construction



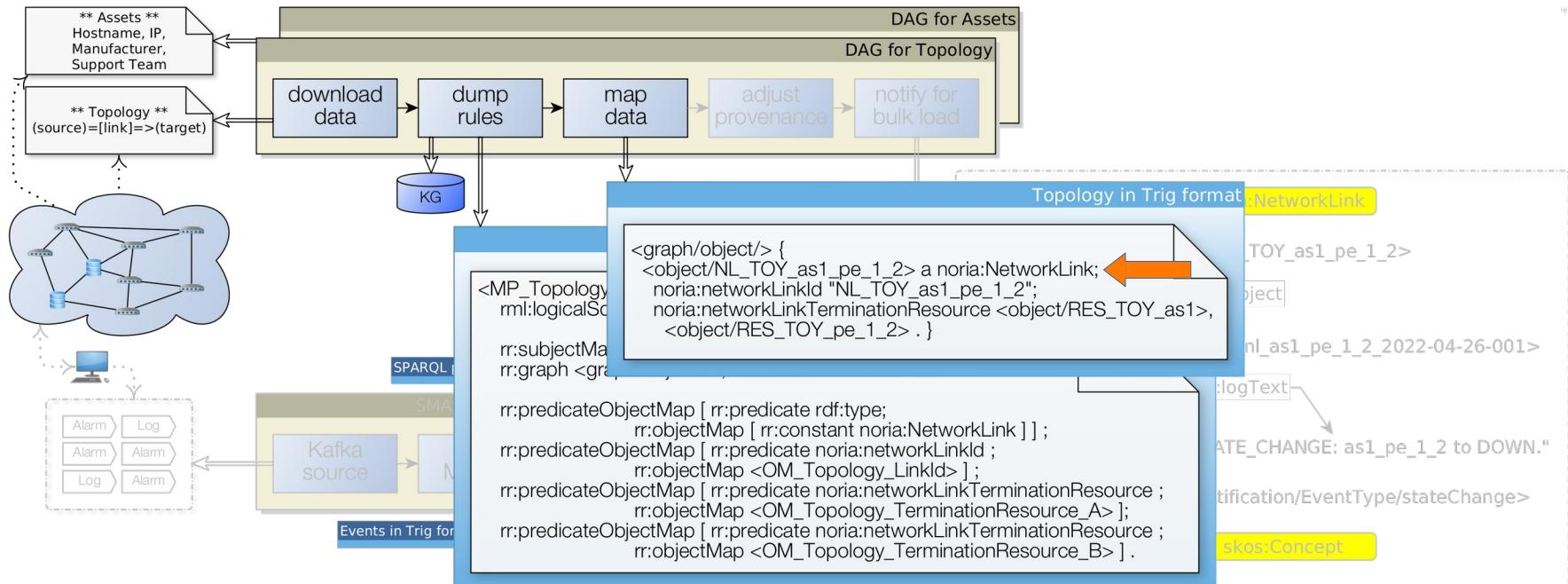
# Knowledge Graph Construction 1/5

Dump RML rules for static data.



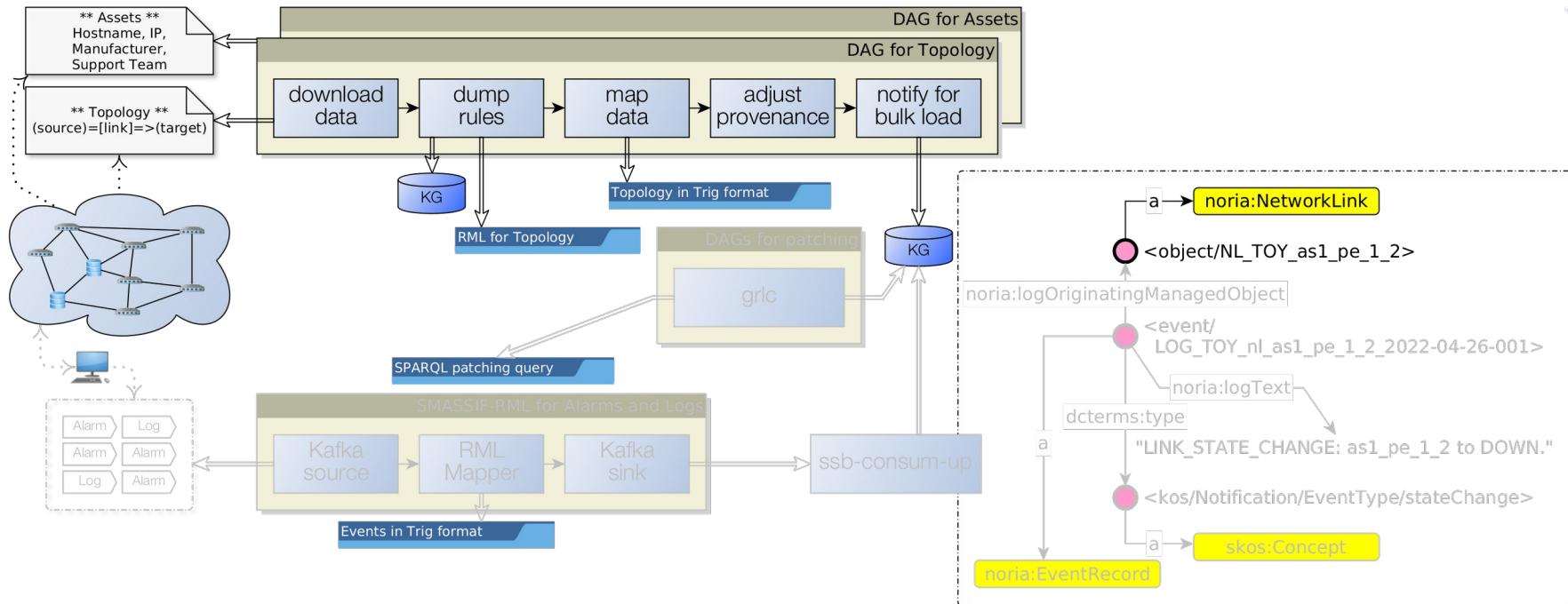
# Knowledge Graph Construction 2/5

Mapping data using RML rules produces triples.



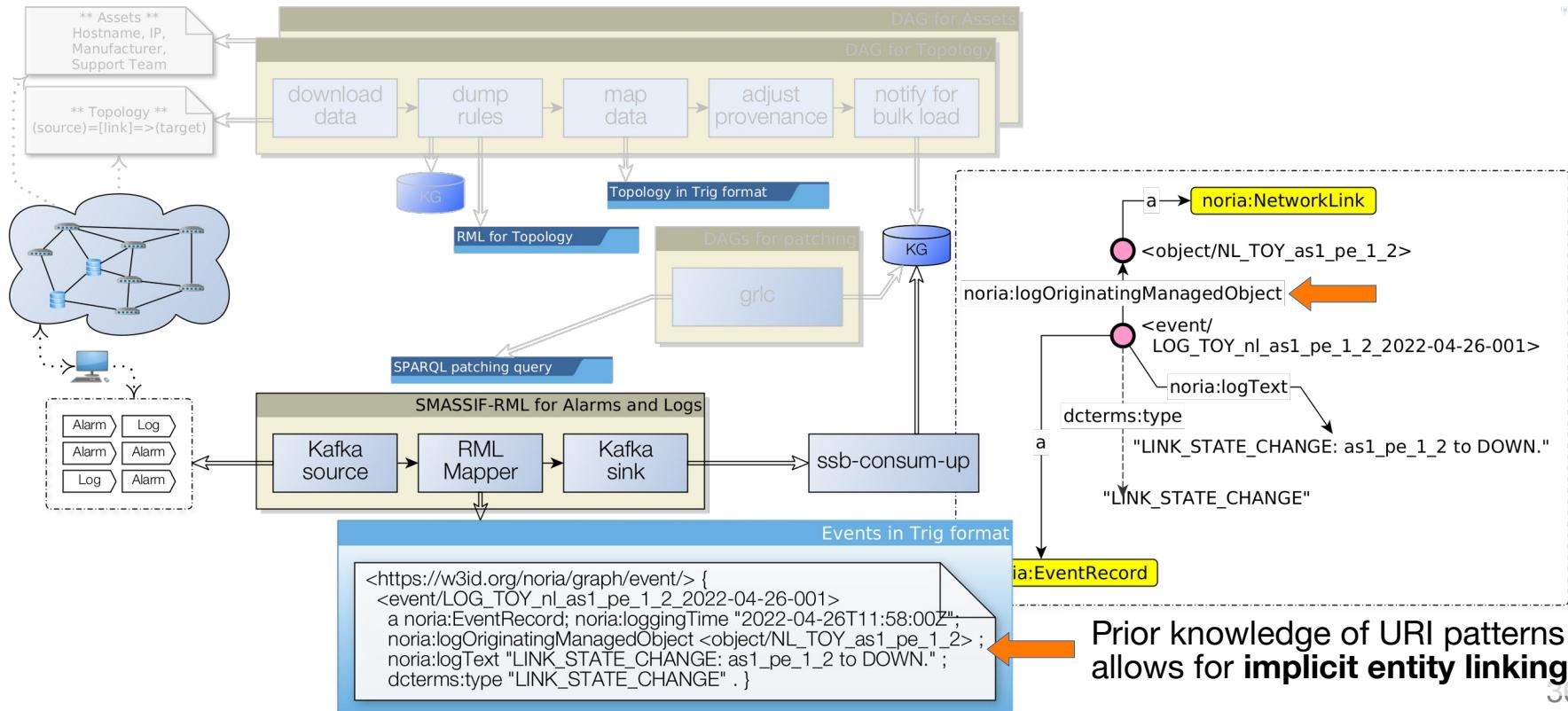
# Knowledge Graph Construction 3/5

Inserting the graph data.



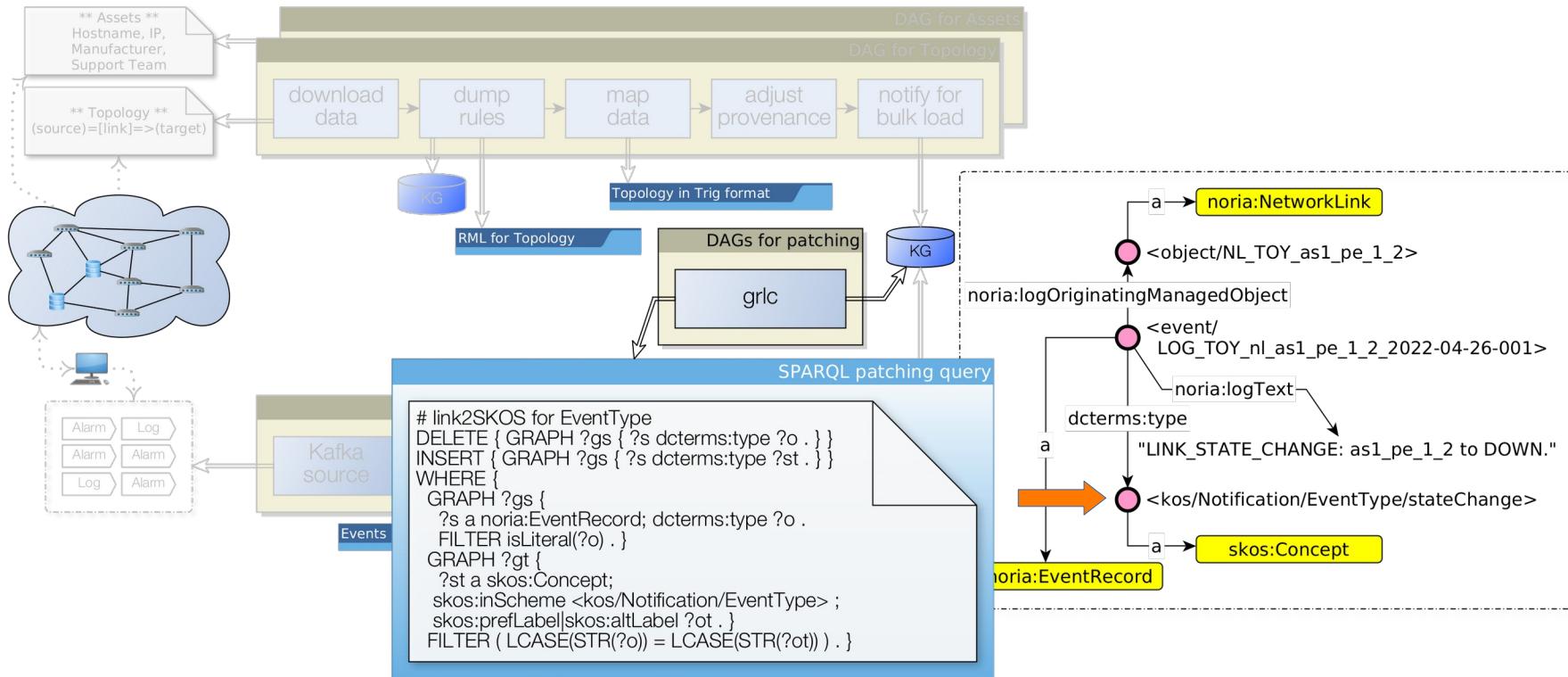
# Knowledge Graph Construction 4/5

Mapping data using RML rules for streamed data and inserting triples in the graph store.

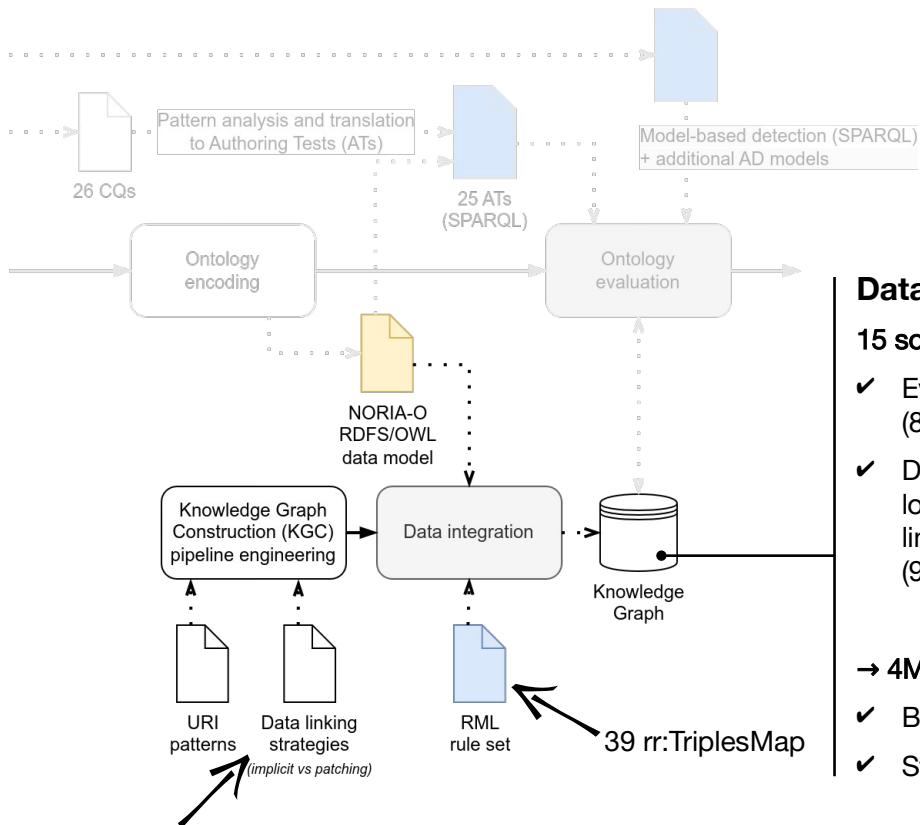


# Knowledge Graph Construction 5/5

Using patching queries for explicit linking of entities.



# Evaluation and Results



42 patching SPARQL queries

- 16 literal2SKOS,
- 19 literal2URI,
- 7 addShortcut.

RQ. 1 - Anomaly model production & utilization with heterogeneous data  
RQ. 2 - Constraints on the internal representation of data and knowledge

# Exploiting the **ICT** systems knowledge

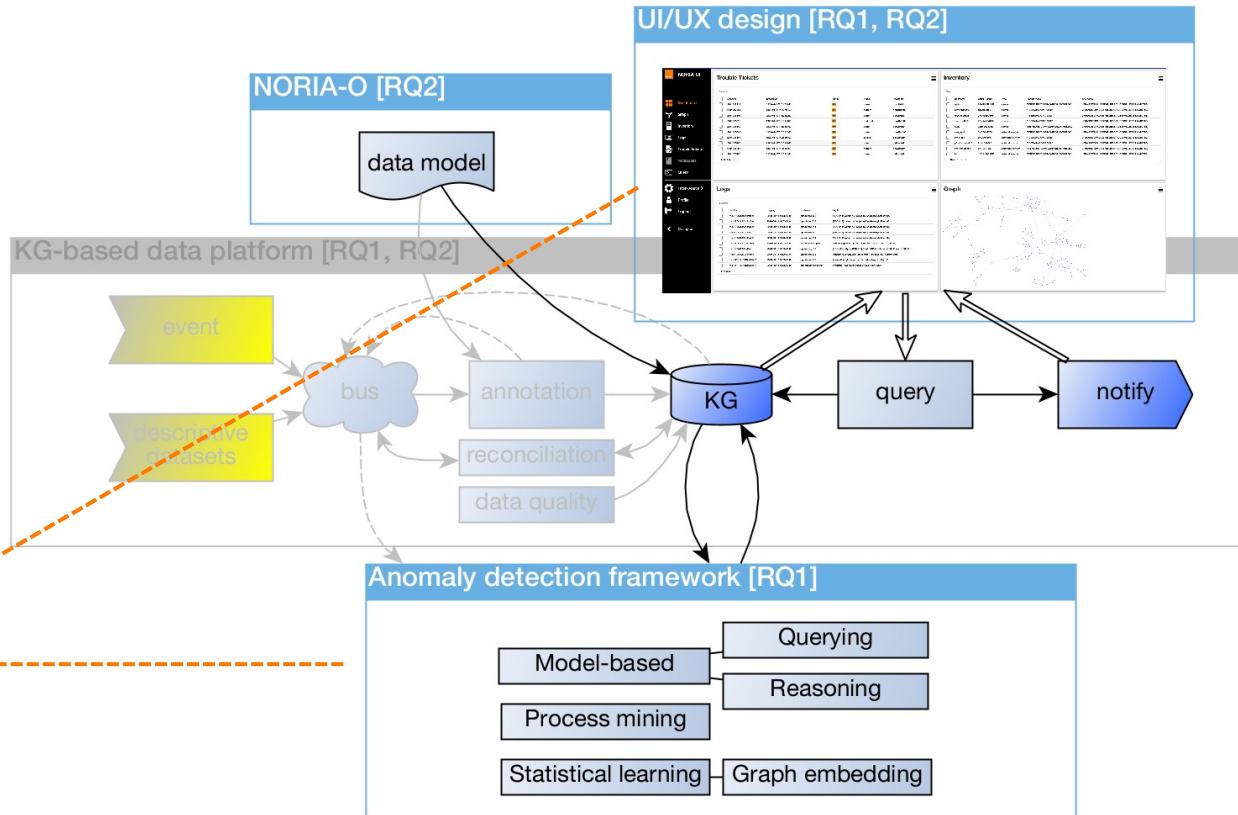
Part II



# I Research Roadmap

Part II

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

# A Cartography of Anomaly Detection Techniques

103 references analyzed: what are the approaches and data structures used, and when are these techniques applied in a business process?

Approach	System Design	Detection & Classification	Diagnostic Aid
Rule-based	1 20,0 %	5 13,2 %	0 0,0 %
Model checking	1 20,0 %	2 5,3 %	1 8,3 %
Knowledge-based	2 <b>40,0 %</b>	6 15,8 %	6 <b>50,0 %</b>
Markov model	0 0,0 %	1 2,6 %	0 0,0 %
Graph-based	1 20,0 %	10 26,3 %	5 41,7 %
ML-based	0 0,0 %	14 <b>36,8 %</b>	0 0,0 %
Overall	5 9,1 %	38 <b>69,1 %</b>	12 21,8 %

❑ Akoglu et al. **Graph-Based Anomaly Detection and Description: A Survey**. Data Mining and Knowledge Discovery, 2015.

❑ Pang et al. **Deep Learning for Anomaly Detection: A Review**. ACM Computing Surveys, 2020.

❑ He et al. **A Survey on Automated Log Analysis for Reliability Engineering**. ACM Computing Surveys, 2021.

❑ González-Granadillo et al. **Security Information and Event Management (SIEM): Analysis, Trends, and Usage in Critical Infrastructures**. Sensors, 2021.

# A Cartography of Anomaly Detection Techniques

103 references analyzed: what are the approaches and data structures used, and when are these techniques applied in a business process?

Approach	System Design	Detection & Classification	Diagnostic Aid
Rule-based	1 20,0 %	5 13,2 %	0 0,0 %
Model checking	1 20,0 %	2 5,3 %	1 8,3 %
Knowledge-based	2 <b>40,0 %</b>	6 15,8 %	6 <b>50,0 %</b>
Markov model	0 0,0 %	1 2,6 %	0 0,0 %
Graph-based	1 20,0 %	10 26,3 %	5 41,7 %
ML-based	0 0,0 %	14 <b>36,8 %</b>	0 0,0 %
Overall	5 9,1 %	38 <b>69,1 %</b>	12 21,8 %

Graph-based approach in all three usage stages: a significant portion of the addressed problems involves the **interconnected nature of the data**.

Prevalence of **logic-based** approaches in the design and diagnostic aid stages, as opposed to **correlation-based** approaches in the detection & classification stage.

55/103 emerged with:

- Primary application domain close to the NetOps and SecOps fields,
- Practicality falling into an **incident management** stage.

Predominance of works applicable to the detection & classification stage.

# A Cartography of Anomaly Detection Techniques

103 references analyzed: what are the approaches and data structures used, and when are these techniques applied in a business process?

Data structures	Approach	System Design	Detection & Classification	Diagnostic Aid
Order relation, e.g. event logs & alarms, network traffic dump, temperature.	Rule-based	1 20,0 %	5 13,2 %	0 0,0 %
Graph (static or streaming), e.g. network topology.	Model checking	1 20,0 %	2 5,3 %	1 8,3 %
Tabular data, e.g. assets with their characteristics.	Knowledge-based	2 <b>40,0 %</b>	6 15,8 %	6 <b>50,0 %</b>
Multi-dimensional data points.	Markov model	0 0,0 %	1 2,6 %	0 0,0 %
Mixed approaches, i.e. combination of the above structures.	Graph-based	1 20,0 %	10 26,3 %	5 41,7 %
	ML-based	0 0,0 %	14 <b>36,8 %</b>	0 0,0 %
	Overall	5 9,1 %	38 <b>69,1 %</b>	12 21,8 %

General tendency for **detection & classification** approaches to focus on the *temporal evolution* of systems, while **diagnostic aid** approaches tend to focus on a broader *context of the system's state*.

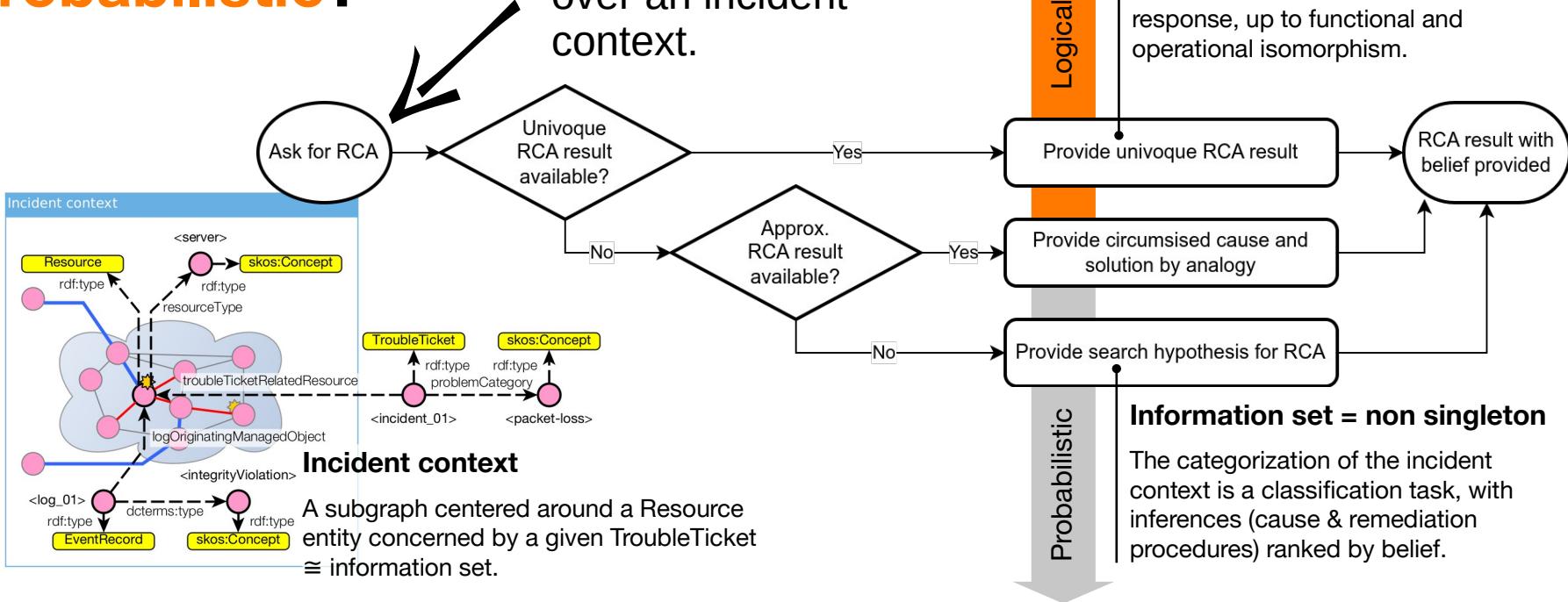
## Challenges in Anomaly Detection (AD)

Potential difficulties in choosing algorithmic methods arise: they individually do not capture and analyze phenomena that involve **temporal, structural, logical, and probabilistic** aspects **simultaneously**.



# Logical or Probabilistic?

Incident management triggers a Root Cause Analysis (RCA) activity over an incident context.



**From logical to probabilistic:** the local network behavior knowledge serves as crisp foundation upon which we can build and combine, up to scale uncertainty and zero-shot diagnosis.

# Synergistic Reasoning

Susie analysing the situation:

- « Is there any pattern in a given set of logs/alarms? » (CQ 9)
- « Which sequence of events led to the incident? » (CQ 12)
- « What past incidents are similar to a given incident? » (CQ 14)



Design choices for AI-based Anomaly Detection

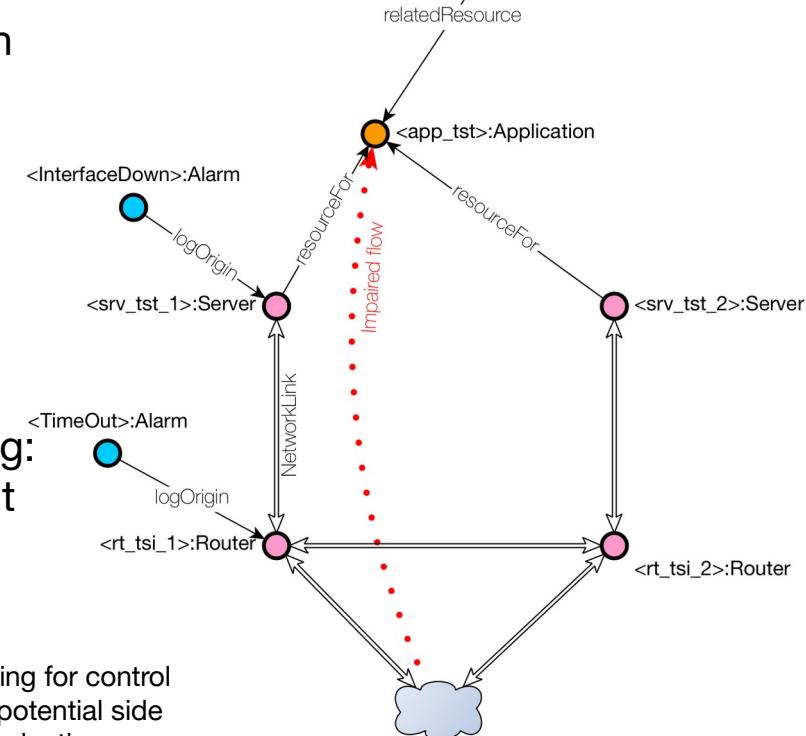
- Logical *vs* probabilistic,
- Single inference model *vs* model stacking.

Why choose? Let's **combine techniques** to leverage their strengths, such as explainability and generalization, and achieve a **broader coverage of detection cases** compared to using a single model.

Design choices for cooperative decision-making:  
**sequential** and/or **auto-organizing** multi-agent decision-making.

## Sequential model combination

An experimental plan that is easier to implement initially, allowing for control over the progression from logical to probabilistic, and limiting potential side effects caused by agent interactions that would necessitate evaluating non-monotonic reasoning, which is more laborious.



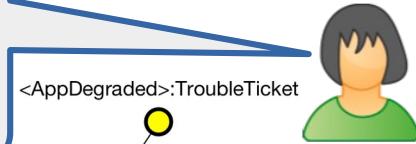
# Synergistic Reasoning

Susie analysing the situation:

« Is there any pattern in a given set of logs/alarms? » (CQ 9)

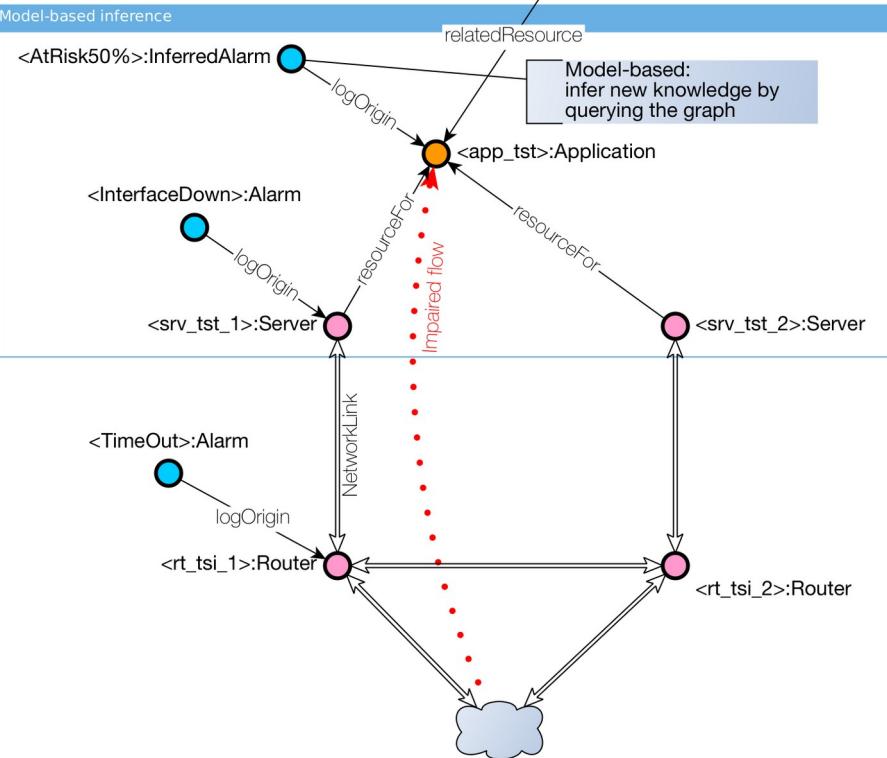
« Which sequence of events led to the incident? » (CQ 12)

« What past incidents are similar to a given incident? » (CQ 14)



**Model-Based Design.** Query the graph to retrieve anomalies and their context

- k out-of n devices with faults
- User with unusual account rights
- Absence of traffic on an interface supposed to be active



# Synergistic Reasoning

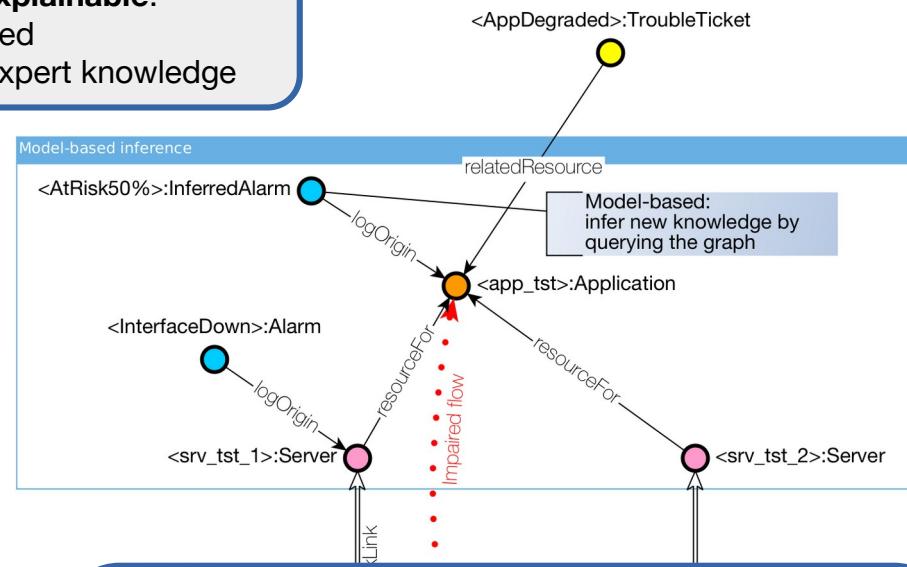
**Model-Based Design.** Query the graph to retrieve anomalies and their context

- k out-of n devices with faults

```
CONSTRUCT {
  ?App noria:atRisk "K out-of N (50%)" . } # <= alerting
WHERE {
  SELECT ?App
    (COUNT(DISTINCT ?Res) AS ?ResTotal)
    (COUNT(DISTINCT ?ResImp) AS ?ResWithImpact)
  WHERE {
    # Get all resources participating in a given
    # application/service ...
    ?Res a noria:Resource ;
      noria:resourceForApplication ?App .
    # Get resources with an alarm, if any ...
    OPTIONAL {
      ?Event a noria:EventLog ;
        noria:eventLogOriginatingManagedObject ?Res .
      BIND (?Res AS ?ResImp) }
    # The k out-of n condition ...
    GROUP BY ?App
    HAVING ( (?ResWithImpact / ?ResTotal) >= 0.5)
  }
```

The query (in SPARQL syntax) is implicitly **explainable**:

- Logic-based
- Reflects expert knowledge



Knowledge mining: **query patterns can be extracted** from the database of operational support systems, up to expert validation. E.g. 12 SPARQL query patterns found by browsing the « incident description » field of a private dataset made of 139 `noria:TroubleTicket` entities.

L. Tailhardat et al. **Leveraging Knowledge Graphs For Classifying Incident Situations in ICT Systems**. ARES'23.

# Synergistic Reasoning

Susie analysing the situation:

« Is there any pattern in a given set of logs/alarms? » (CQ 9)

« Which sequence of events led to the incident? » (CQ 12)

« What past incidents are similar to a given incident? » (CQ 14)

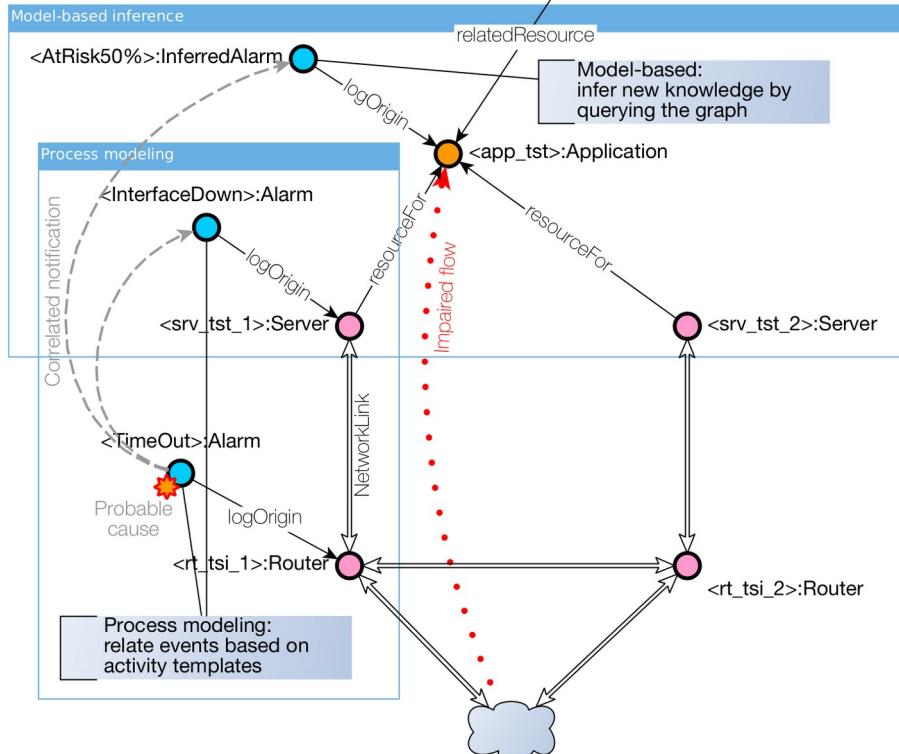


**Model-Based Design.** Query the graph to retrieve anomalies and their context

- k out-of n devices with faults
- User with unusual account rights
- Absence of traffic on an interface supposed to be active

**Process mining.** Align a sequence of entities to activity models, then use this relatedness to guide the repair

- (EnergyLoss)=>(TimeoutAlert)=>(LossOfSignal)
- (LoginFail)=>(LoginFail)=>(LoginFail)

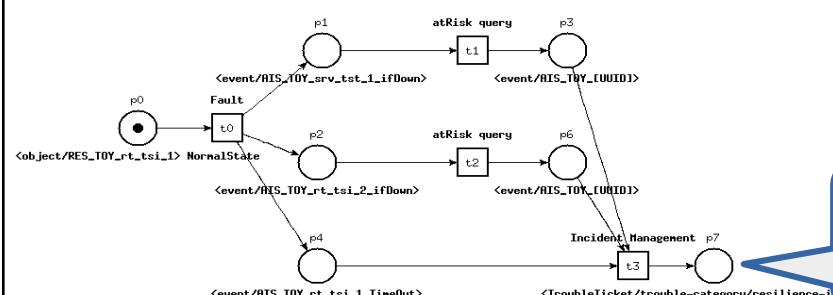


# Synergistic Reasoning

- Model-Based Design.** Query the system to retrieve anomalies and their context
- k out-of n devices with faults
  - User with unusual account rights
  - Absence of traffic on an interface supposed to be active

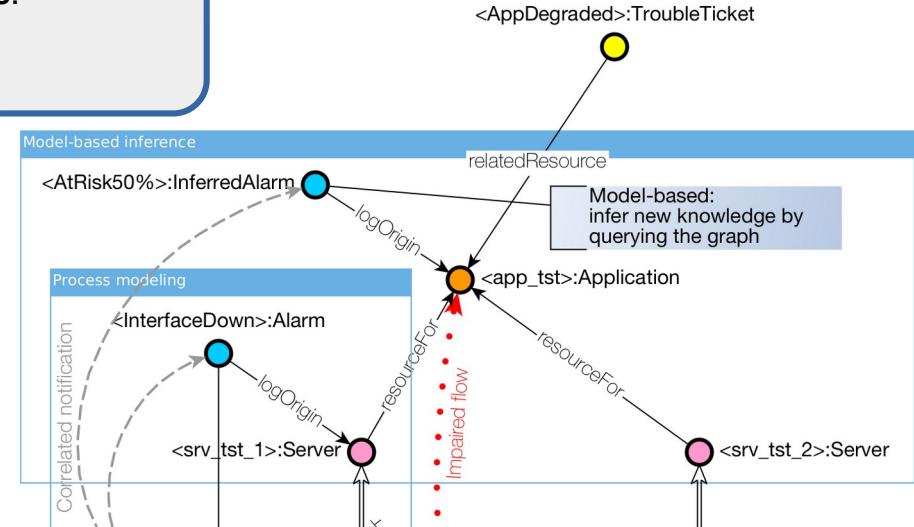
**Process mining.** Align a sequence of entities to activity models, then use this relatedness to guide the repair

- **(EnergyLoss)=>(TimeoutAlert)=>(LossOfSignal)**



Procedural models, e.g. in Petri net form, are also implicitly **explainable**:

- Logic-based
- Reflect expert knowledge



Knowledge mining: **procedural models can be extracted** too, up to expert refinement and validation...

# Synergistic Reasoning

Susie analysing the situation:

« Is there any pattern in a given set of logs/alarms? » (CQ 9)

« Which sequence of events led to the incident? » (CQ 12)

« What past incidents are similar to a given incident? » (CQ 14)

**Model-Based Design.** Query the graph to retrieve anomalies and their context

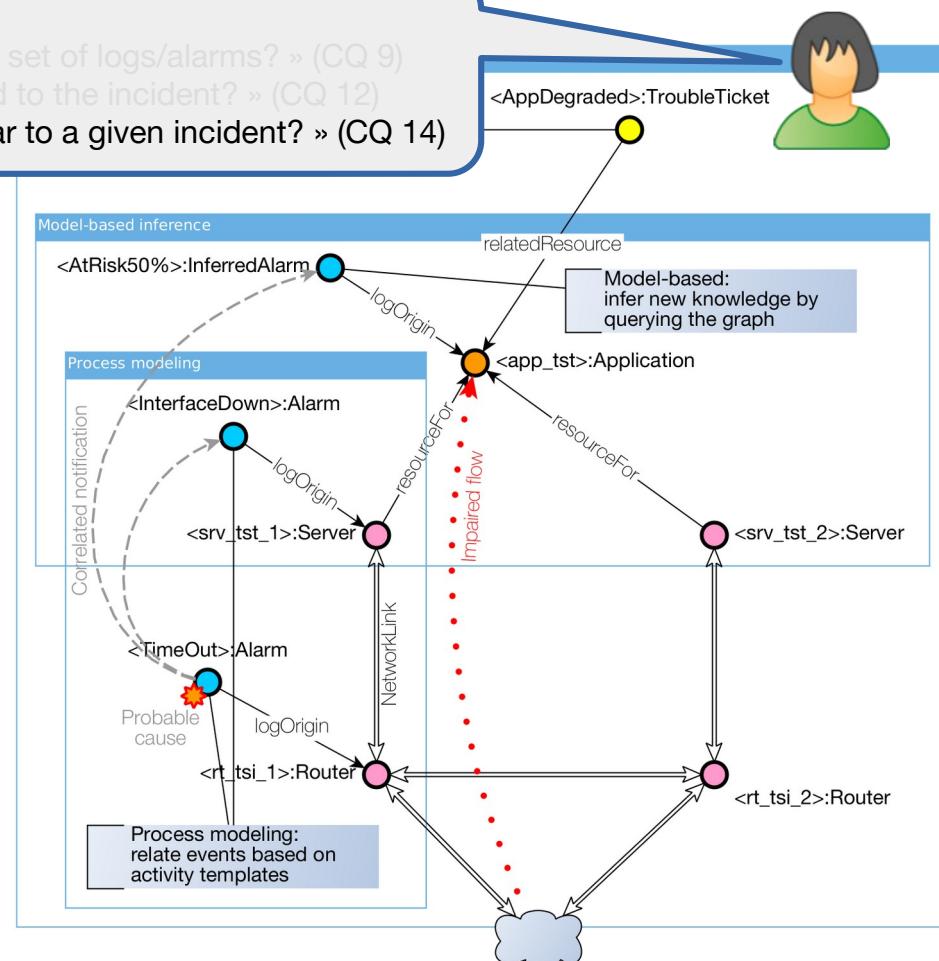
- k out-of n devices with faults
- User with unusual account rights
- Absence of traffic on an interface supposed to be active

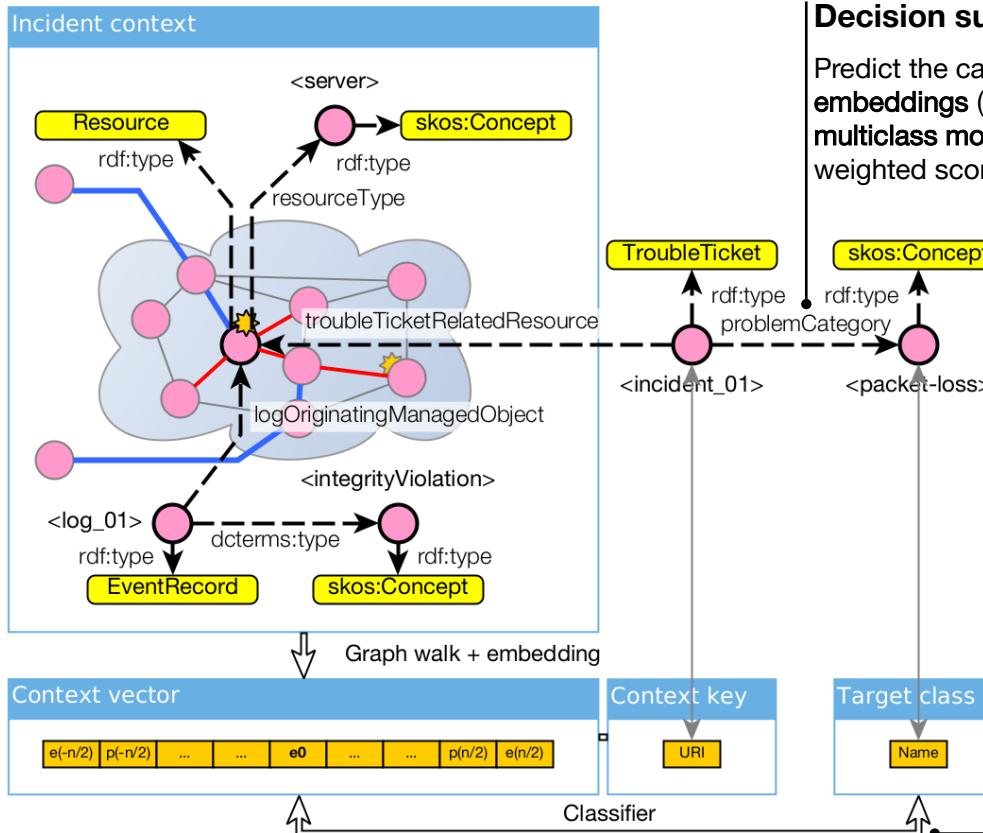
**Process mining.** Align a sequence of entities to activity models, then use this relatedness to guide the repair

- (EnergyLoss)=>(TimeoutAlert)=>(LossOfSignal)
- (LoginFail)=>(LoginFail)=>(LoginFail)

**Statistical Learning.** Relate entities based on context similarities, then use this relatedness to alert and guide the repair

- The hidden cause of the trouble ticket on server 1 is a “data leak” attack that started on server 2





## Decision support = classification problem

Predict the category of a trouble ticket using **graph embeddings** (random walk + CBOW model) and a **multiclass monolabel classifier** (random forest, F1 weighted score model selection).

## Evaluation & results

Dataset from the knowledge graph construction pipeline:

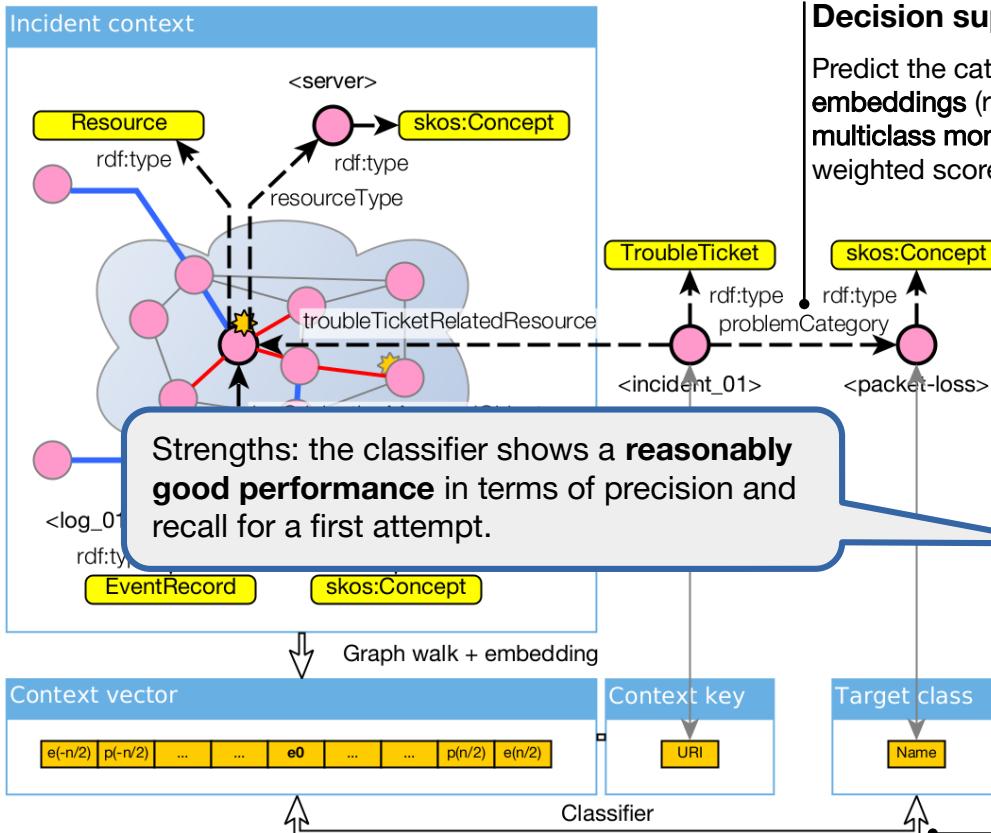
- 15 sources → 4M triples (400K entities)
- 138 noria:TroubleTicket entities
- 5 target class (noria:troubleTicketCategory property)

Best model shows **0.81 F1 weighted score**:

- Supervised learning, 75/25 % stratified fixed-split dataset
  - ✓ Interrupted service: 77 entities (55.8%), 0.97 w. F1
  - ✓ Degraded QoS: 22 (15.9%), 0.75
  - ✓ No service impact: 22 (15.9%), 0.62
  - ✓ Defect to be qualified: 13 (9.4%), 0.57
  - ✓ Equipment failure: 4 (2.9%), 0.00
- Embeddings with walk depth = 8, walk count = 30
- Random forest with max tree depth = 5, tree count = 20



# Statistical Learning



## Decision support = classification problem

Predict the category of a trouble ticket based on its context. Use **graph embeddings** (random walk + GAT) + **multiclass monolabel classification** + weighted score model selection.

Caveats: the **dataset is too small** (for some classes in particular) + available context for trouble ticket entities is **not systematically consistent**.

## Evaluation & results

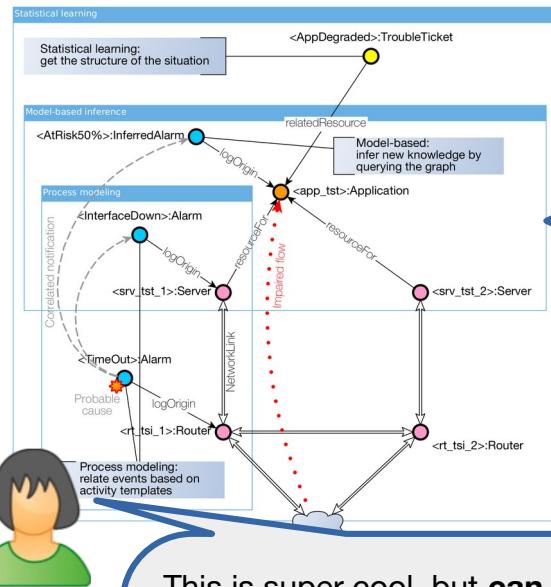
Dataset from the knowledge graph construction pipeline:

- 15 sources  $\rightarrow$  4M triples (400K entities)
- 138 noria:TroubleTicket entities
- 5 target class (noria:troubleTicketCategory property)

Best model shows **0.81 F1 weighted score**:

- Supervised learning, 75/25 % stratified fixed-split dataset
  - ✓ Interrupted service: 77 entities (55.8%), 0.97 w. F1
  - ✓ Degraded QoS: 22 (15.9%), 0.75
  - ✓ No service impact: 22 (15.9%), 0.62
  - ✓ Defect to be qualified: 13 (9.4%), 0.57
  - ✓ Equipment failure: 4 (2.9%), 0.00
- Embeddings with walk depth = 8, walk count = 30
- Random forest with max tree depth = 5, tree count = 20





This is super cool, but **can we make it simple**, considering that I have Service Level Agreements (SLAs) to respect ?

By « we », I mean incident managers, network supervision experts, cybersecurity analysts, system architects, etc.

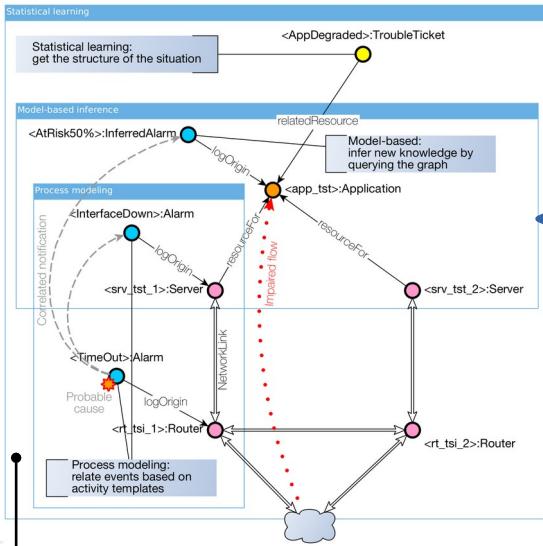
ID	Created at	Status	Priority	Urgency	Intervention difficult	Intervention details	Description	Actions
AM011409	06/09/2023, 08:00:00	Open	Low	Medium	Medium	Next intervention reporting issue...	Customer call reporting issue...	<input type="button" value="View"/> <input type="button" value="Edit"/> <input type="button" value="Delete"/>
AM000005	06/09/2023, 05:00:00	Open	Medium	High	Medium	Next intervention reporting issue...	Customer call reporting issue...	<input type="button" value="View"/> <input type="button" value="Edit"/> <input type="button" value="Delete"/>
AM000006	06/09/2023, 06:00:00	Open	Medium	Medium	Medium	Next intervention reporting issue...	Customer call reporting issue...	<input type="button" value="View"/> <input type="button" value="Edit"/> <input type="button" value="Delete"/>

## UI/UX design (co-design with Orange operation experts)

- ✓ Development, deployment and evaluation of a Web-based client-server architecture leveraging a knowledge graph structured by NORIA-O.
- ✓ Principle: providing access to information about the network's life based on four complementary facets derived from the knowledge graph.

L. Tailhardat et al. **NORIA UI: Efficient Incident Management on Large-Scale ICT Systems Represented as Knowledge Graphs.** ARES'24.

# Evaluation and Results



The screenshot shows the NORIA UI interface, which includes the following components:

- Trouble Tickets:** A table listing trouble tickets with columns for ID, Created at, Status, Priority, Urgency, Intervention attitude, Description, and Actions. Examples include "TOY\_TOT\_RTS\_00000000000000000000000000000000" and "TOY\_TOT\_RTS\_00000000000000000000000000000000".
- Events and alarms:** A table listing events and alarms with columns for ID, Logging time, Severity, Description, Related resource, and Actions. Examples include "Dummy alarm" and "Interface Opened...".
- Resources and applications:** A table listing resources and applications with columns for ID, Entity Name, Type, Installation Date, and Actions. Examples include "AM014409 Application" and "AM000005 Application".
- Graph:** A network graph visualization showing connections between various entities represented by nodes and edges.

## UI/UX design [RQ. 2]

- ✓ **UI/UX evaluation campaign:** 1 month duration, 10 active beta testers, average SUS score = 68.4, correlation of the respondents' profile with the acceptability level (from good to high).

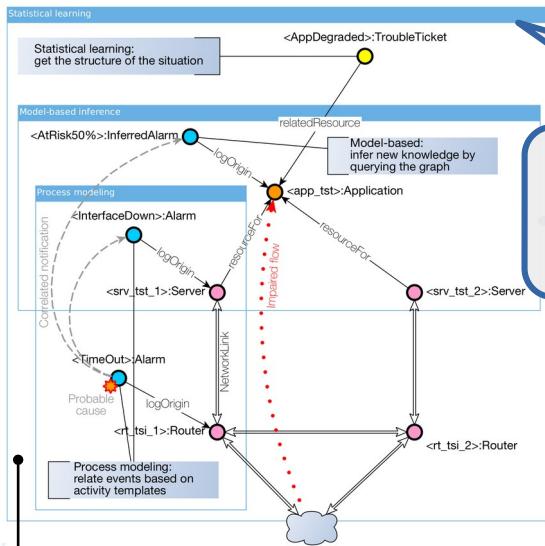
└ L. Tailhardat et al. **NORIA UI: Efficient Incident Management on Large-Scale ICT Systems Represented as Knowledge Graphs.** ARES'24.

RQ. 1 - Anomaly model production & utilization with heterogeneous data  
 RQ. 2 - Constraints on the internal representation of data and knowledge

## Anomaly detection framework leveraging the synergistic reasoning principle [RQ. 1]

- ✓ **Model-based:** 2 SPARQL-based detection cases, 2 reasoning-based cases, and 12 query patterns.
- ✓ **Process mining:** 2 alignment-based detection cases and a Web extension to learn user-network behavioral models.
- ✓ **Statistical learning:** graph-embedding-based classifier achieving an interesting 0.81 F1 score as an initial attempt.

# Evaluation and Results

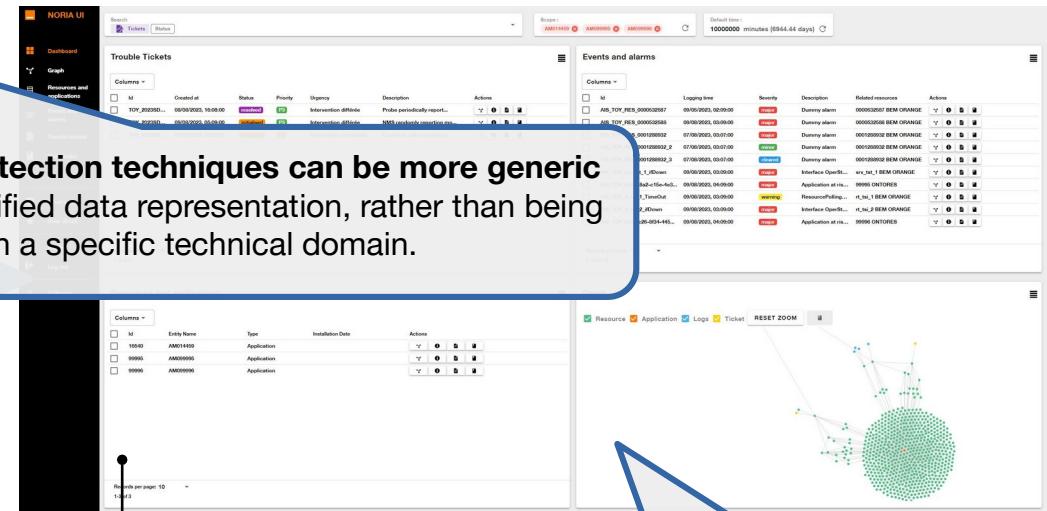


**Anomaly detection framework leveraging the synergistic reasoning principle [RQ. 1]**

- ✓ Model-based: 2 SPARQL-based detection cases, 2 reasoning-based cases, and 12 query patterns

**Cooperative decision-making:** each detection 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.

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



**UI/UX design [RQ. 2]**

- ✓ UI/UX evaluation campaign: 1 month duration, 10 active beta testers, average age 35, gender 50% male, 50% female, with diverse professional profiles.
- Facilitating knowledge graph use without specific training is achieved by **linearizing the exploration process**, and implementing **tailored interaction mechanisms** for incident management.

RQ. 1 - Anomaly model production & utilization with heterogeneous data  
RQ. 2 - Constraints on the internal representation of data and knowledge

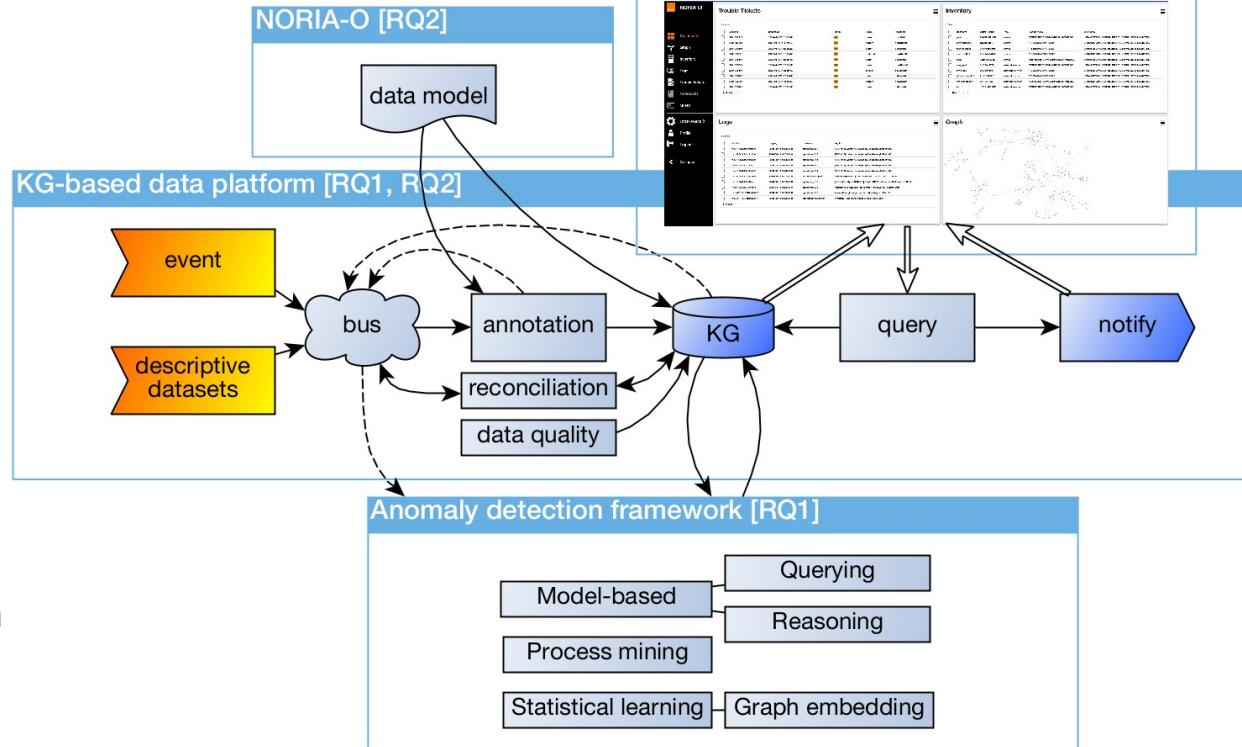
# Anomaly Detection using Knowledge Graphs and Synergistic Reasoning

Conclusion



# 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 :

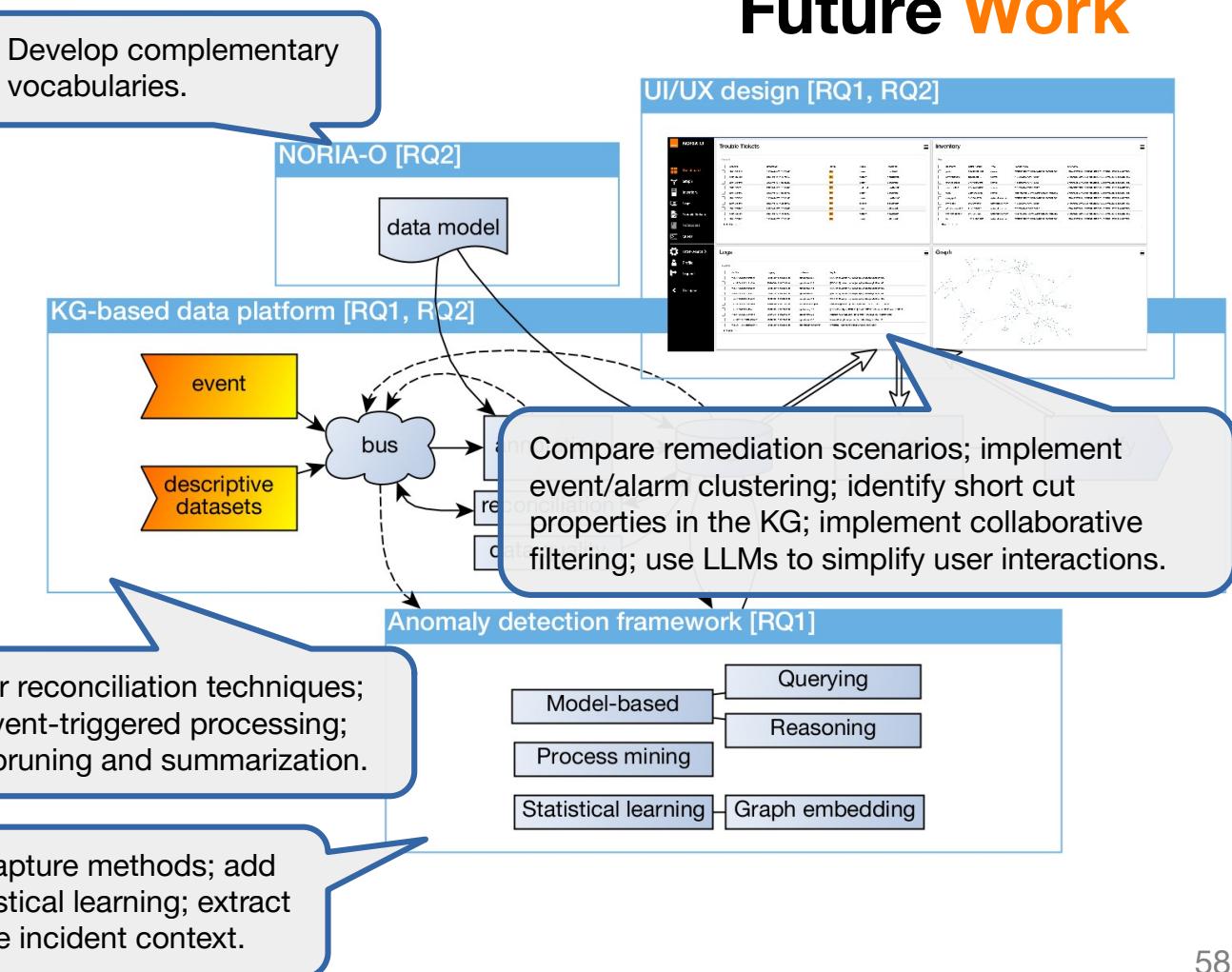
- Achieve **cross technical domain anomaly detection** with intrinsic explainability and probabilistic reasoning capabilities.
- Identify and share strengths and weaknesses of infrastructures (FMEA).

RQ. 1 - Anomaly model production & utilization with heterogeneous data  
RQ. 2 - Constraints on the internal representation of data and knowledge

# Future Work

Towards new subjects:

- Knowledge Graphs at the company scale.
- Neuro-symbolic multi-agent system for synergistic reasoning.
- Root cause analysis with graph generation and causal models.
- Cybersecurity risk assessment and moving target defense.

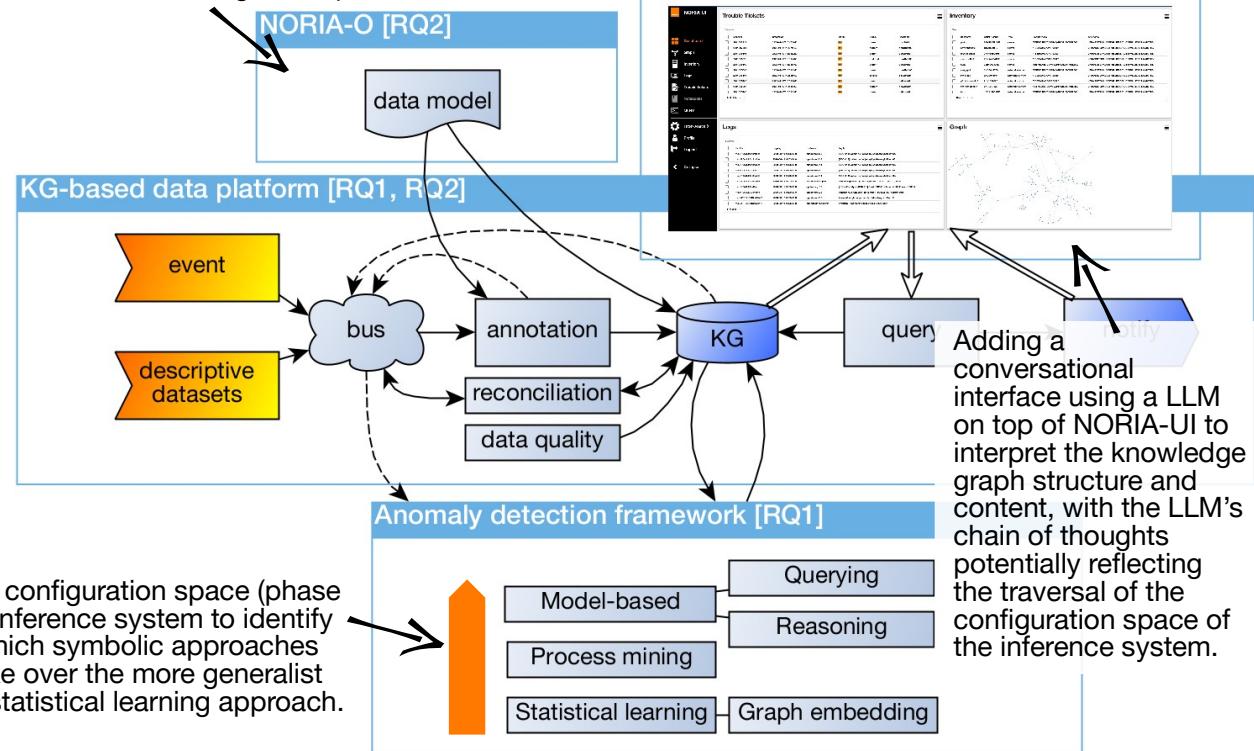


# Future Work

Towards new subjects:

- Knowledge Graphs at the company scale.
- Neuro-symbolic multi-agent system for synergistic reasoning.
- Root cause analysis with graph generation and causal models.
- Cybersecurity risk assessment and moving target defense.

Using Competency Questions as guides for selecting an approach, either individually or in a sequence reflecting the incident management process.



Traversing the configuration space (phase space) of the inference system to identify the point at which symbolic approaches definitively take over the more generalist nature of the statistical learning approach.

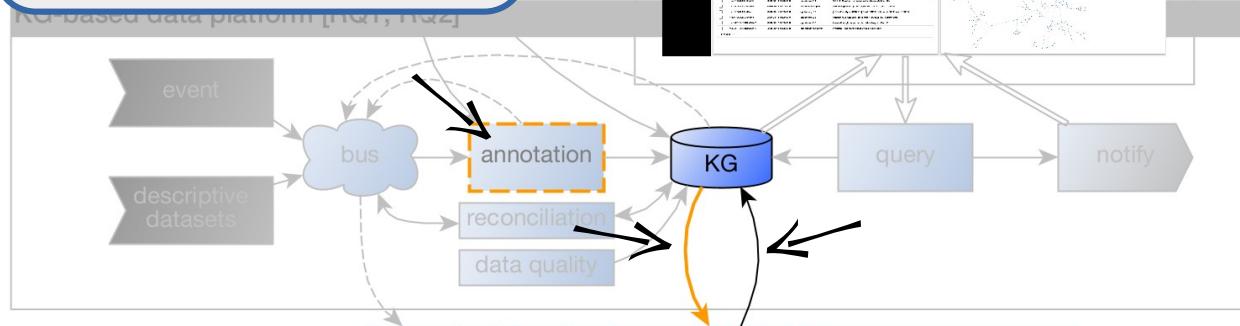
**How to select and ideally order each anomaly detection approach to ensure trustworthy decision-making?**

# Closing Thoughts for HYSNC -

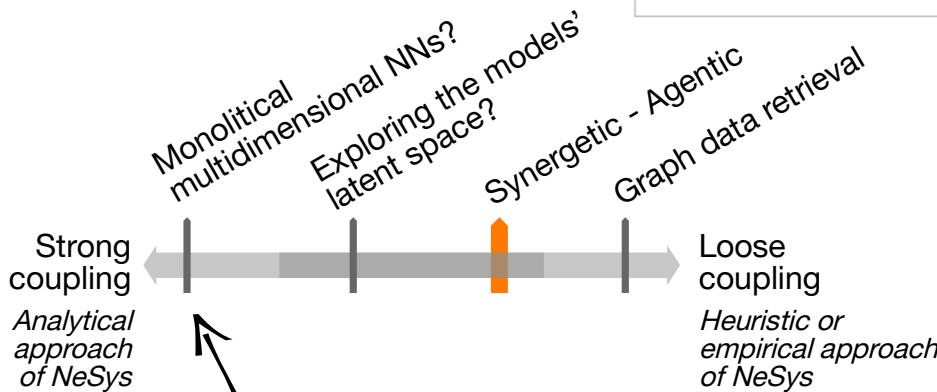
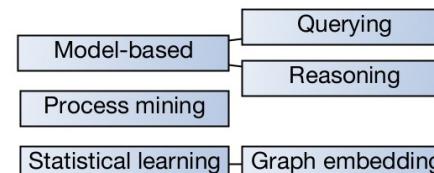
## Current trends in NeSys and Knowledge Graphs (opinion)

- R&D professionals are primarily looking into loosely connected KG and NN parts.
- RDF/RDFS/OWL KG practitioners overlook automated reasoning, favoring heterogenous data management capabilities of the SemWeb stack.

Artefacts of the old challenges of combining logic with probabilities? Lack of specific math skills in projects? Need for a new paradigm?



## Anomaly detection framework [RQ1]



Going analytic: we still have not solved the challenge of a **monolithic NeSys model**, **if that is indeed the goal to be achieved**.

# Additional materials

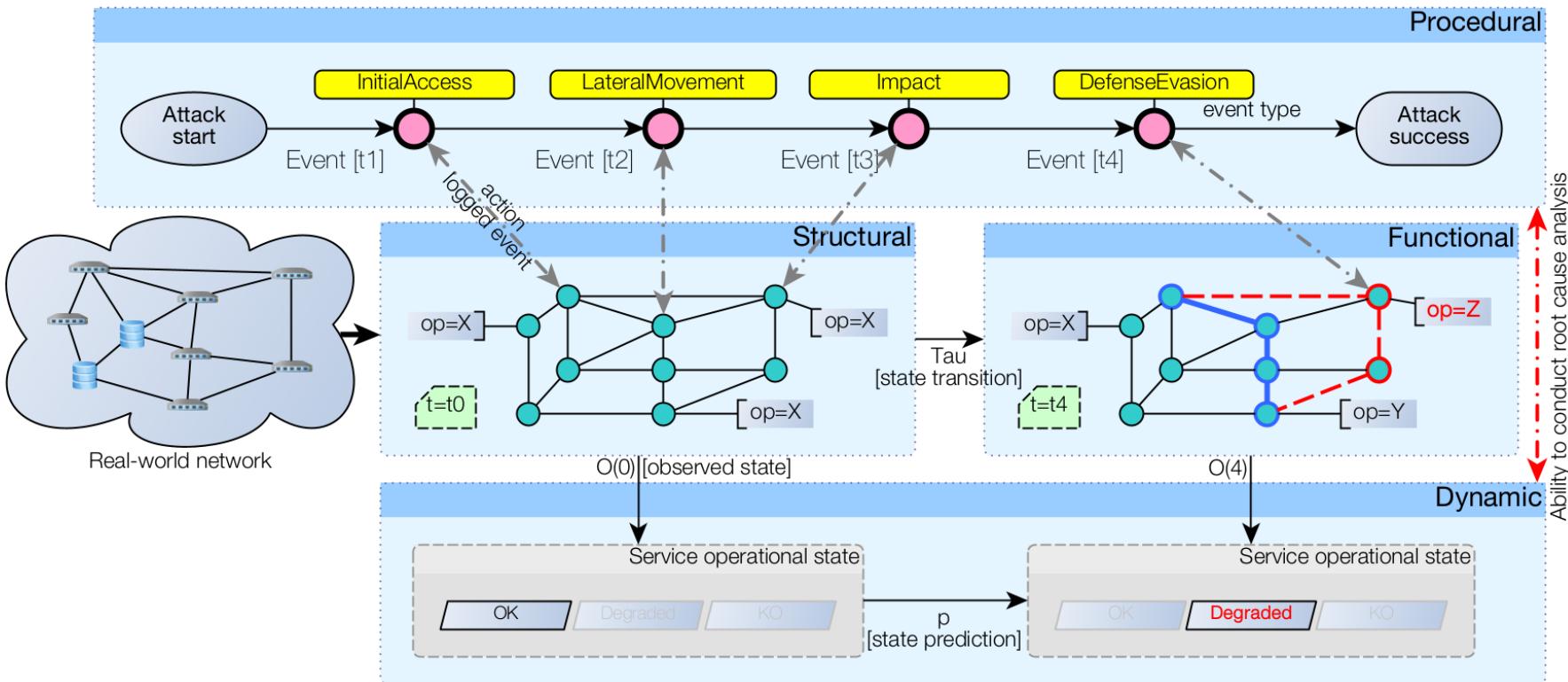
Appendix



# Code and Dataset

- **NORIA-O**, an RDF data model for IT networks, events and operations information.  
<https://w3id.org/noria>
- **grlc**, a fork of CLARIAH/grlc with SPARQL UPDATE and GitLab interface features.  
<https://github.com/Orange-OpenSource/grlc>
- **SMASSIF-RML**, a Semantic Web stream processing solution with declarative data mapping capability based on a modified version of the RMLMapper-java tool and extensions to the StreamingMASSIF framework.  
<https://github.com/Orange-OpenSource/SMASSIF-RML>
- **ssb-consum-up**, a Kafka to SPARQL gateway enabling end-to-end Semantic Web data flow architecture with a Semantic Service Bus (SSB) approach.  
<https://github.com/Orange-OpenSource/ssb-consum-up>
- **SemNIDS**, bringing semantics into Network Intrusion Detection Systems.  
<https://github.com/D2KLab/SemNIDS>
- **Dynagraph**, network dumping and Web app for live 3D graph rendering of streamed graph data derived from traces.  
<https://github.com/Orange-OpenSource/dynagraph>
- **Graphameleon**, a Web extension that captures Web navigation traces and transforms them into a RDF graph for further exploration.  
<https://github.com/Orange-OpenSource/graphameleon>
- **Graphameleon dataset**, an RDF dataset of Web navigation traces, generated by the Graphameleon Web extension.  
<https://github.com/Orange-OpenSource/graphameleon-ds>
- **LLM4KE**, a dataset of RDF data models, and code for generating competency questions.  
<https://github.com/D2KLab/llm4ke>

# ICT System State Transition Model



The representation of a network can be divided into four facets: **structural**, **functional** (the blue path indicates an operational data flow, the red path a faulty flow), **dynamic**, and **procedural** (logged events are related to cyber-security attack tactics from the MITRE ATT&CK matrix).  $Tau$  stands for state transition,  $O(t)$  for observed state at time  $t$ , and  $p$  for state prediction.

The 26 NORIA-O competency questions,  
available at <https://w3id.org/noria/cqs/>

# NORIA-O Competency Questions 1/3

1. Which resource/application/site is concerned by a given incident?
2. What assets are shared by a given asset chain?
3. What logs and alarms are coming from a specified resource?
4. Which metrics are coming from a specified resource?
5. To which event family does this log belong and is this event normal or abnormal?
6. What events are associated with a given event?
7. Which agent/event/resource caused the event under analysis?
8. What do the various fields in the log refer to?
9. Is there any pattern in a given set of logs/alarms?
- 10.What interventions were carried out on this resource that could have caused the incident?
- 11.What was the root cause of the incident?
- 12.Which sequence of events led to the incident?
- 13.On which resource did this sequence of events take place and in which order?
- 14.What past incidents are similar to a given incident?

# NORIA-O Competency Questions 2/3

The 26 NORIA-O competency questions, available at <https://w3id.org/noria/cqs/>

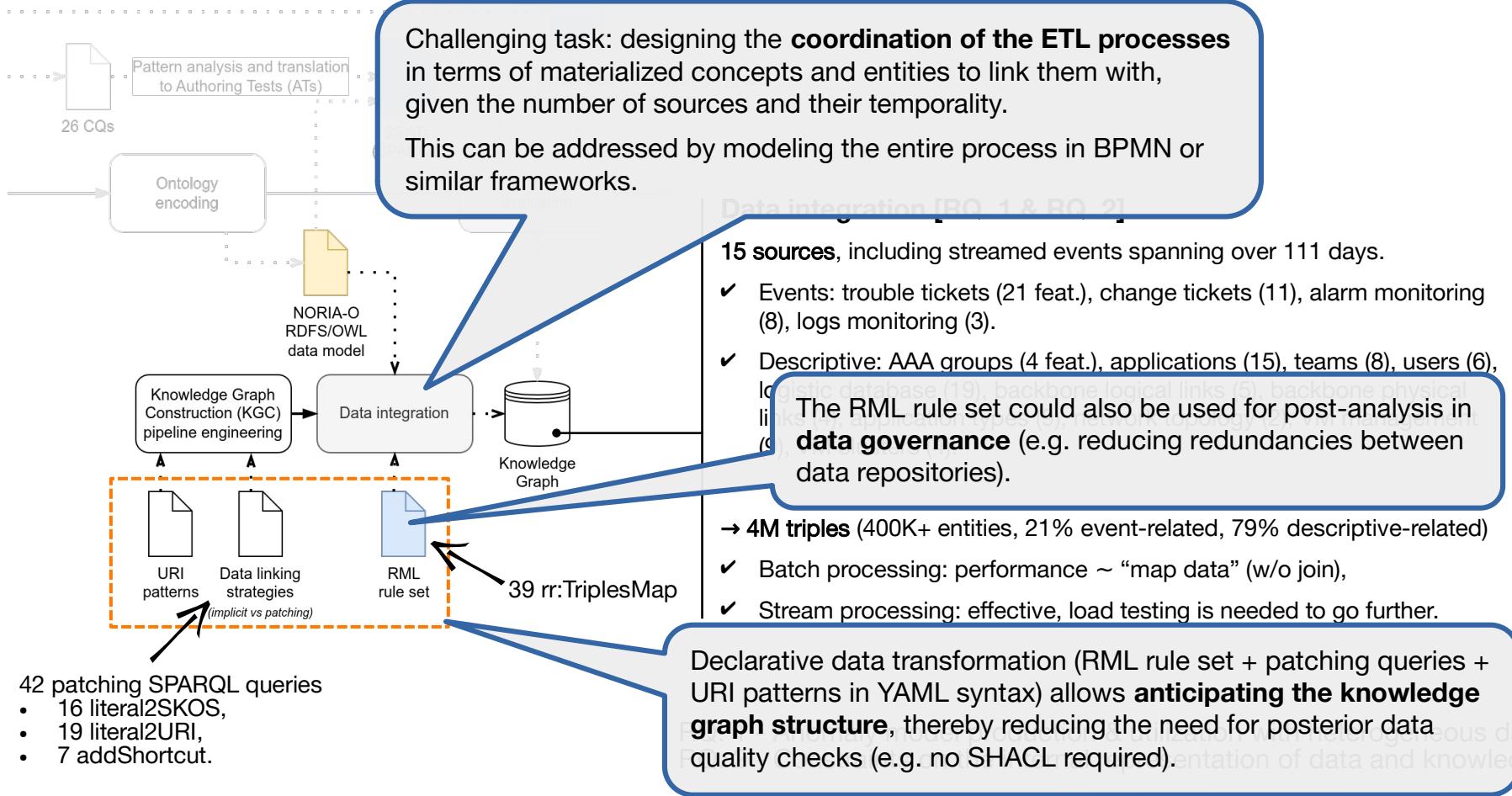
- 15.What operation plan (automation, operating procedures, etc.) could help us solve the incident?
- 16.What corrective actions have been carried out so far for a given incident?
- 17.What is the list of actions taken that led to the resolution of the incident?
- 18.Given all the corrective actions carried out so far for the incident, what assumptions covered the actions taken?
- 19.What has been the effect of the corrective actions taken so far for the incident?
- 20.Given all the corrective actions carried out so far for the incident, what possible actions could we still take?
- 21.What is the summary of this incident and its resolution?
- 22.Which agents were involved in the resolution of the incident?
- 23.What is the financial cost of this incident if it occurs?
- 24.How long before this incident is resolved?
- 25.What are the vulnerabilities and the associated risk levels of this infrastructure?
- 26.What is the most likely sequence of actions that would cause this infrastructure to fail?

# NORIA-O Competency Questions 3/3

St.	Fu.	Dy.	Pr.	Competency Questions
✓	✓			What assets are shared by a given asset chain?
✓		✓		Which entity (resource/application/site) is concerned by a given incident?
✓		✓		On which resource did this sequence of events take place and in which order?
		✓		What corrective actions have been carried out so far for a given incident (who, what, where)?
		✓		What interventions were carried out on this resource that could have caused the incident?
		✓		What operation plan (automations, operating procedures, etc.) could help us solve the incident?
		✓		Given all the corrective actions carried out so far for the incident, what possible actions could we still take?

The four knowledge facets to represent (St.: structural, Fu.: functional, Dy.: dynamic, Pr.: procedural) map to a subset of NORIA-O competency questions.

# NORIA-DI Evaluation and Results



# KGC Dataset Example

```
{  
  "id": "TOY2022TT",  
  "creationDateTime": "2022-04-26T11:58:00Z",  
  "description": "Toy example: service access  
    Failure from term1. Probable cause: network issue.",  
  "detectionDateTime": "2022-04-26T11:58:00Z",  
  "lastUpdate": "2022-04-26T12:07:00Z",  
  "isNotificationEnable": false,  
  "category": { "label": "Impaired service" },  
  "priority": { "label": "P2" },  
  "status": [  
    {  
      "code": "InProgress",  
      "isCurrentStatus": true,  
    },  
    "troubleTicketCharacteristic": [...],  
    "note": [  
      {  
        "text": "Service access diagnosis: no route to  
          srv1.",  
        "recordingDate": "2022-04-26T12:05:00Z",  
        "author": "LF001",  
        "operationType": { "label": "Comment" }  
      }, [...]  
    ]  
  ]  
}
```

JSON

```
<https://w3id.org/noria/document/TT_TOY2022TT>  
  a noria:TroubleTicket;  
  dcterms:created "2022-04-26T12:00:00Z";  
  dcterms:description """Toy example: service  
    access failure from term1. Probable cause:  
    Network issue.""";  
  dcterms:identifier "TOY2022TT";  
  dcterms:modified "2022-04-26T12:07:00Z";  
  dcterms:extent "P0Y0M0DT0H10M0S" ;  
  noria:troubleTicketDetectionDateTime  
    "2022-04-26T11:58:00Z";  
  noria:troubleTicketRelatedResource  
    <https://w3id.org/noria/object/RES_TOY_term1>;  
  noria:troubleTicketStatusCurrent  
    <https://w3id.org/noria/ontology/kos/  
      TroubleTicket/status/current> ;  
  noria:documentStatusHistory  
    <https://w3id.org/noria/event/  
      LOG_TOY_TT_TOY2022TT_STATUS_Current> ;  
  dcterms:hasPart  
    <https://w3id.org/noria/document/  
      TTN_TOY2022TT_2022-04-26T12:05:00Z CU_LF001>,  
    <https://w3id.org/noria/document/  
      TTN_TOY2022TT_2022-04-26T12:07:00Z CU_LF004>;
```

Turtle

TroubleTicket (raw and Turtle syntax): excerpt from the NORIA-O dataset, available at  
<https://w3id.org/noria/>

List of use cases from expert panel interviews, in simplified form.

# Incident Diagnosis

## Activity Cases

1. Circumscribe assets and causes search space for multi-applications incident situations
2. Alert on impaired service situations occurring on (distributed) fail-over architectures
3. Assess legitimacy of a given network flow
4. Track single identity from a set of various activity traces
5. Analyze false-positive and recurrent cyber security alerts
6. Analyze compliance of web navigation traces from institutional website

# Data Structures and Algorithmic Methods

Approach	Seq. data	Seq. data (network)	Time series	<i>Ordered</i> (1,2,3)	Graph	Graph streams	Tabular	Data points	Mixed seq.+ graph	Mixed seq.+ tab.	Mixed seq.+ unstr.	<i>Mixed</i> (9,10,11)	
	[%]	[%]	[%]	$\Sigma$	[%]	[%]	[%]	[%]	[%]	[%]	[%]	[%]	$\Sigma$ [%]
<b>Design</b>													
G.-based	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	1 10,0	0,0	0,0	1	8,3
K.-based	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	1 10,0	1 <b>100,0</b>	0,0	2	16,7
M. check.	1 7,1	0,0	0,0	<i>1</i> 4,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
R.-based	0,0	0,0	0,0	0,0	1 9,1	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
<b>Detection &amp; Classification</b>													
G.-based	2 14,3	0,0	1 16,7	3 12,0	3 27,3	1 <b>50,0</b>	2 <b>66,7</b>	0,0	1 10,0	0,0	0,0	1	8,3
K.-based	2 14,3	1 20,0	0,0	3 12,0	3 27,3	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Markov	1 7,1	0,0	0,0	<i>1</i> 4,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
ML-based	5 <b>35,7</b>	1 20,0	5 <b>83,3</b>	11 <b>44,0</b>	0,0	1 <b>50,0</b>	0,0	2 <b>100,0</b>	0,0	0,0	0,0	0,0	0,0
M. check.	1 7,1	0,0	0,0	<i>1</i> 4,0	1 9,1	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
R.-based	1 7,1	3 <b>60,0</b>	0,0	<i>4</i> 16,0	1 9,1	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
<b>Diagnostic Aid</b>													
G.-based	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	5 <b>50,0</b>	0,0	0,0	5	<b>41,7</b>
K.-based	0,0	0,0	0,0	0,0	2 18,2	0,0	1 33,3	0,0	2 20,0	0,0	1 <b>100,0</b>	3	25,0
M. check.	1 7,1	0,0	0,0	<i>1</i> 4,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Overall	14 <b>25,5</b>	5 9,1	6 10,9	25 <b>45,5</b>	11 20,0	2 3,6	3 5,5	2 3,6	10 18,2	1 1,8	1 1,8	12	21,8

Distribution (in number and proportion) of the main data structures used within the algorithmic solutions in the analyzed papers, based on the algorithmic approach family and the stage of the incident management process involved. Values in bold highlight the most representative approach for a given data structure. The columns in italics represent cumulative values (ordered = columns 1 + 2 + 3, mixed = columns 9 + 10 + 11) to provide a summary view of similar structures.

# Anomaly Modeling Technique Families

Principles	Strengths	Weaknesses
<b>Model-Based Design</b>		
Query the graph to retrieve anomalies and their context.	Detecting anomalies “recorded” somehow in the graph thanks to the alarm system; straightforward translation of simple anomaly detection rules; multiple abstraction levels (subsumption).	Relyes on expert knowledge; lack of probabilistic reasoning; hard to represent sequential decisions; may require to infer more prior information about the anomaly, e.g. its type using classification.
<b>Process Mining</b>		
Align a sequence of entities to activity models, then use this relatedness to guide the repair.	Detecting anomalies with multiple alerting signals and sequential decisions; replayable models.	Relyes on expert knowledge; may require denoising models; probabilistic relatedness.
<b>Statistical Learning</b>		
Relate entities based on context similarities, then use this relatedness to alert and guide the repair.	Detecting anomalies with multiple alerting signals.	Requires fine tuning of the context definition depending on use case and temporality requirements; probabilistic relatedness.

# The solution-oriented bias

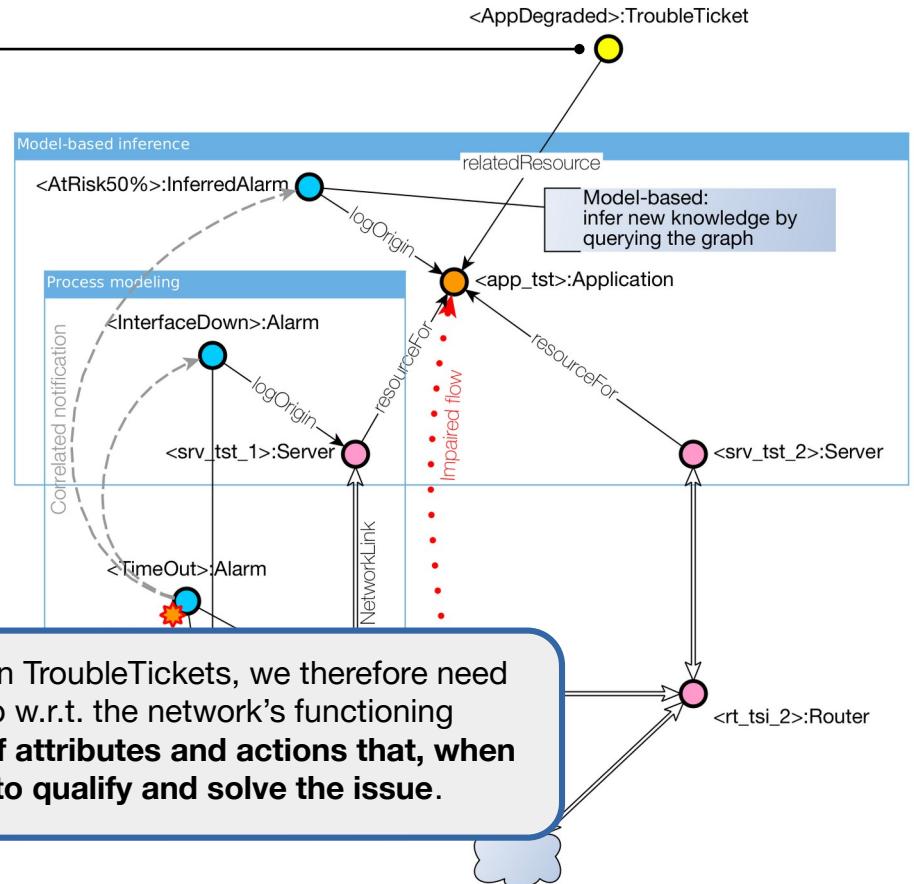
## Need to learn (or deduce) what not to do

TroubleTicket database mining leads to learning a solution-to-undesirable-states-driven mapping function :

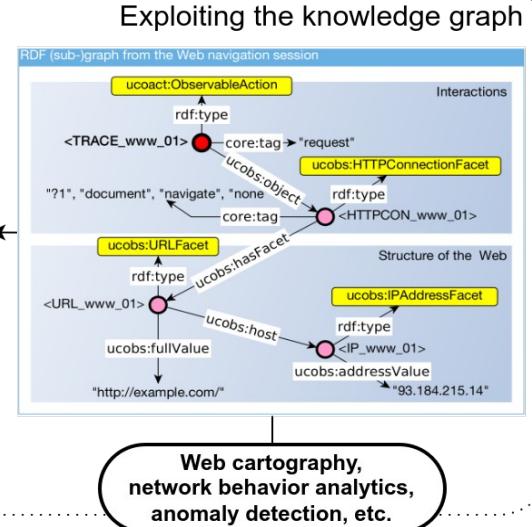
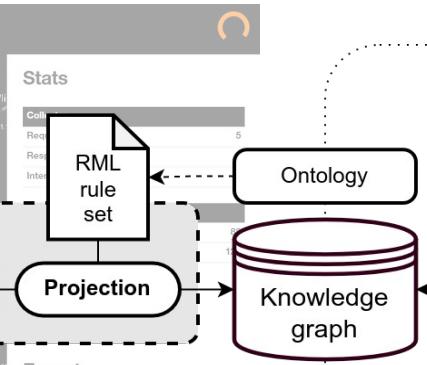
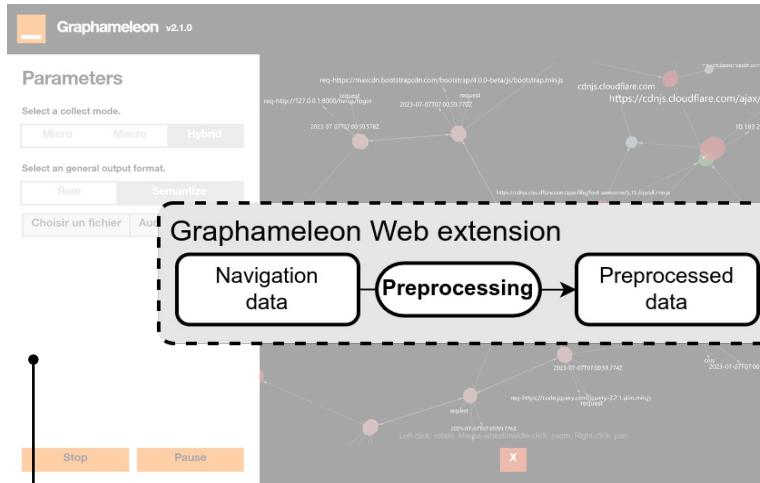
- ✗ Trouble Tickets primarily refer to an incident context and the remediation actions taken, rather than to instances when the network is behaving well.
- ✗ The solution-oriented data is an ill-situation for supervised AI approaches as they require to have evenly distributed class instances for proper classification tasks.

Tackling the **solution-oriented bias** involves **counterfactual reasoning**, i.e. reasoning on events that did not occur but that may have under defined conditions.

Because we cannot rely « only » on TroubleTickets, we therefore need to learn (or deduce) what not to do w.r.t. the network's functioning logic and vulnerabilities: **the set of attributes and actions that, when observed or done, do not allow to qualify and solve the issue.**



# Process Mining



## Collecting procedural models to establish a baseline

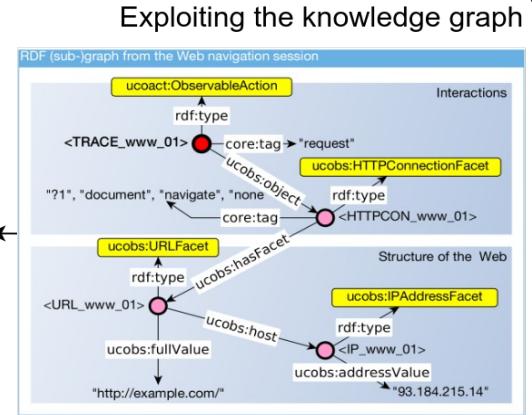
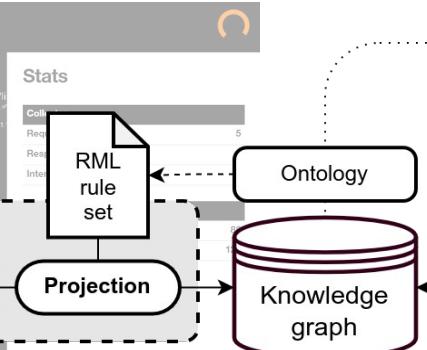
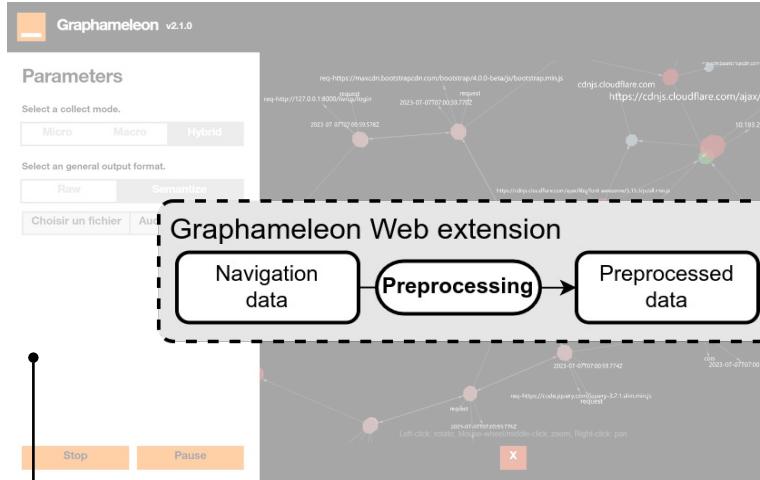
- A Web extension for the **live capture at the browser level** of network requests & user interactions.
- Output of a RDF Knowledge Graph using the UCO ontology.
- Mining procedural models with **process discovery** techniques (PM4Py), and detecting anomalous behaviors with **conformance checking** techniques (PM4Py).

📄 L. Tailhardat et al. **Walks in Cyberspace: Improving Web Browsing and Network Activity Analysis with 3D Live Graph Rendering.** TWC'22.

📄 L. Tailhardat et al. **Graphameleon: Relational Learning and Anomaly Detection on Web Navigation Traces Captured as Knowledge Graphs.** TWC'24.

📄 L. Tailhardat et al. **Graphamélon : apprentissage des relations et détection d'anomalies sur les traces de navigation Web capturées sous forme de graphes de connaissances.** PFIA'24.

# Process Mining



## Collecting procedural models

- A Web extension for the live capture of interactions.
- Output of a RDF Knowledge Graph using the UCO ontology.
- Mining procedural models with **process discovery** techniques (PM4Py), and detecting anomalous behaviors with **conformance checking** techniques (PM4Py).

Procedural models only **capture local processes**, i.e. not the full incident context.

☞ L. Tailhardat et al. **Walks in Cyberspace: Improving Web Browsing and Network Analysis with 3D Live Graph Rendering**. TWC'22.

☞ L. Tailhardat et al. **Graphameleon: Relational Learning and Anomaly Detection Navigation Traces Captured as Knowledge Graphs**. TWC'24.

☞ L. Tailhardat et al. **Graphamélon : apprentissage des relations et détection d'anomalies sur les traces de navigation Web capturées sous forme de graphes de connaissances**. PFIA'24.

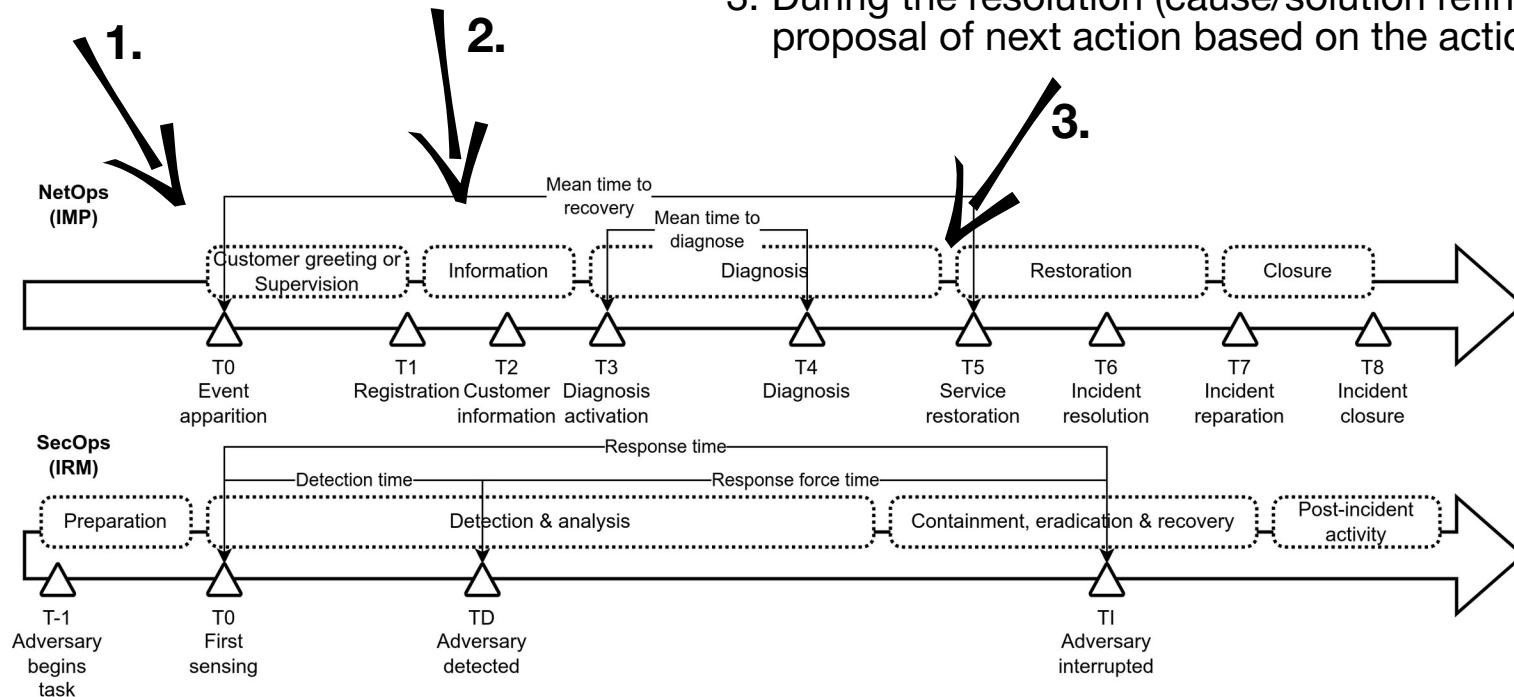
**Web cartography, network behavior analytics, anomaly detection, etc.**

Threshold-based anomaly detection using **model alignment** with observational data may miss micro changes that are important.

# Reasoning Services for Decision Support 1/2

Stages of the incident management process where a recommendation system can be useful:

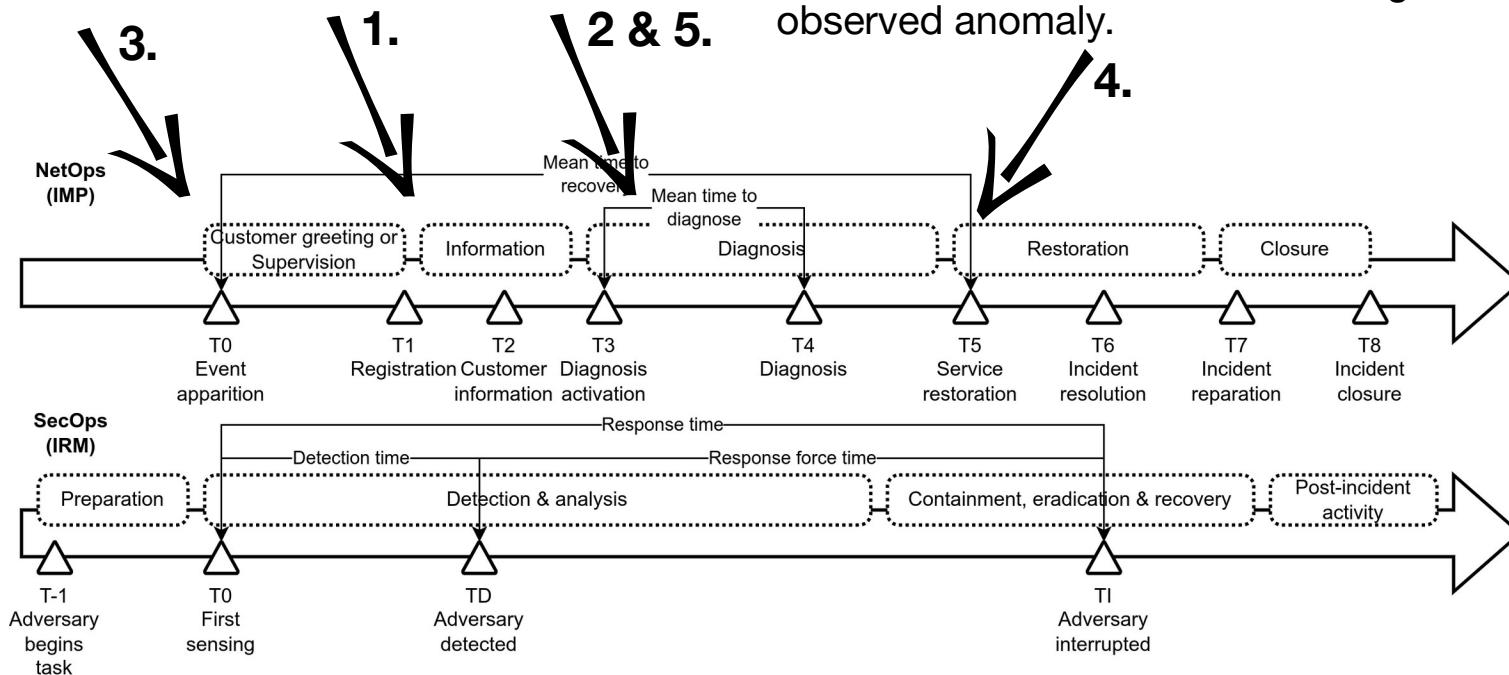
1. Before the ticket creation (early detection),
2. At the ticket opening (cause/solution similarity based on ticket descriptors and context),
3. During the resolution (cause/solution refinement and proposal of next action based on the actions taken).



# Reasoning Services for Decision Support 2/2

## Reasoning services (proposal):

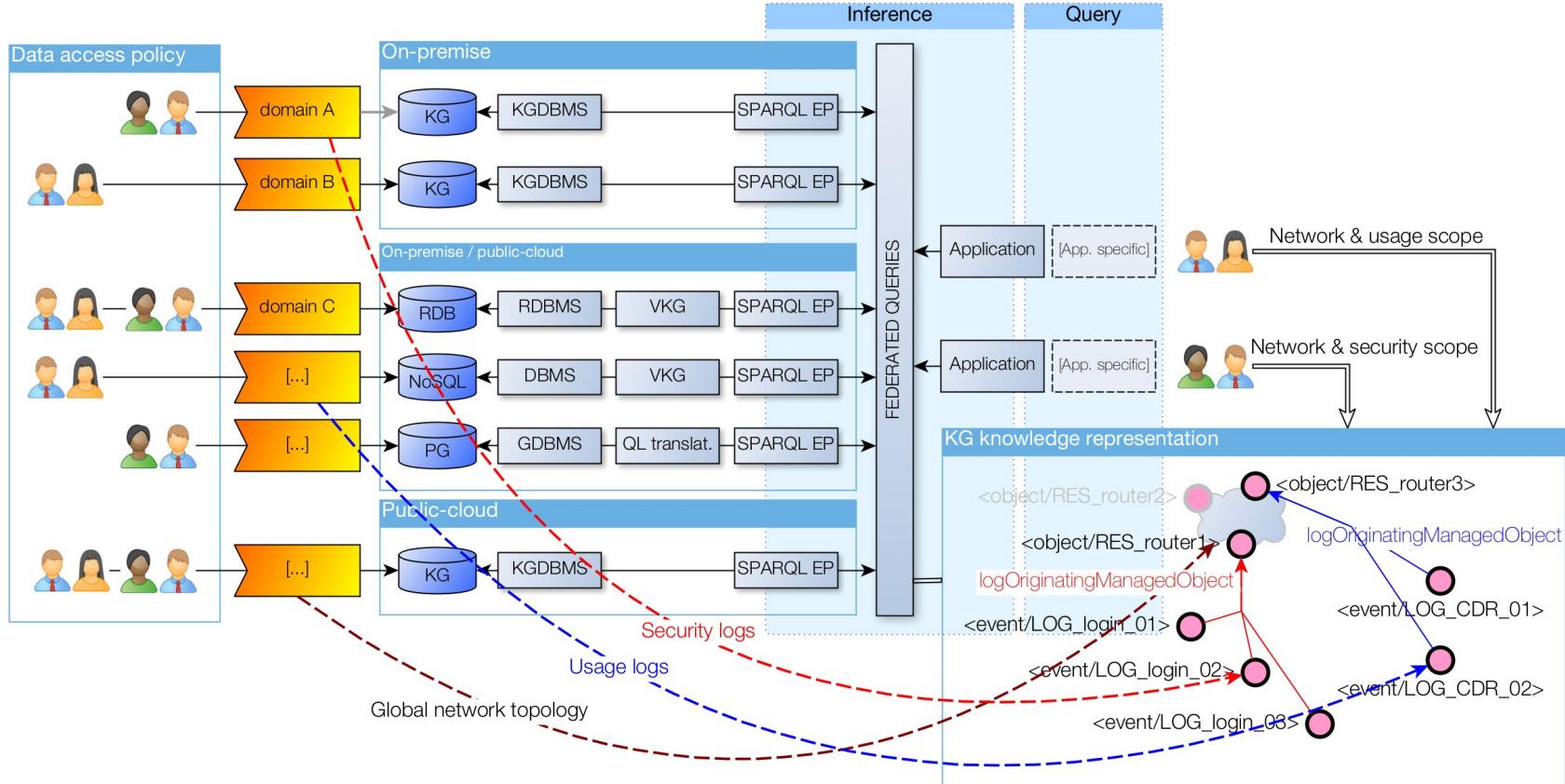
1. Predicting the category of a trouble ticket,
2. Predicting the probable cause of a trouble ticket,
3. Detecting anomalies before a trouble ticket is even created,
4. Adding comments to a given trouble ticket (e.g. next best action to undertake),
5. Calculate the n closest anomalies given an observed anomaly.



# Federating Partitioned Data

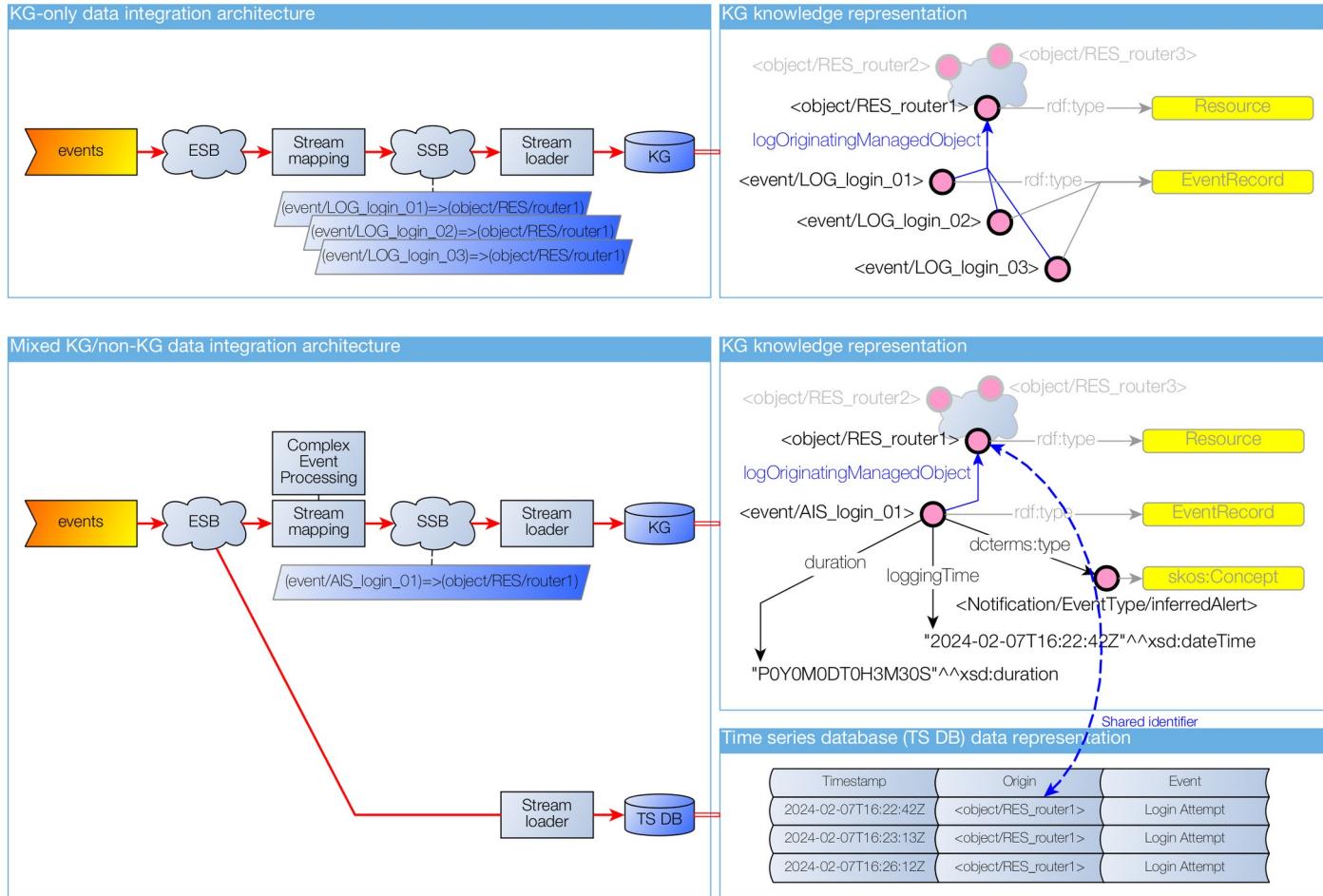
**Federated queries** for providing,

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



# Scaling with Streams

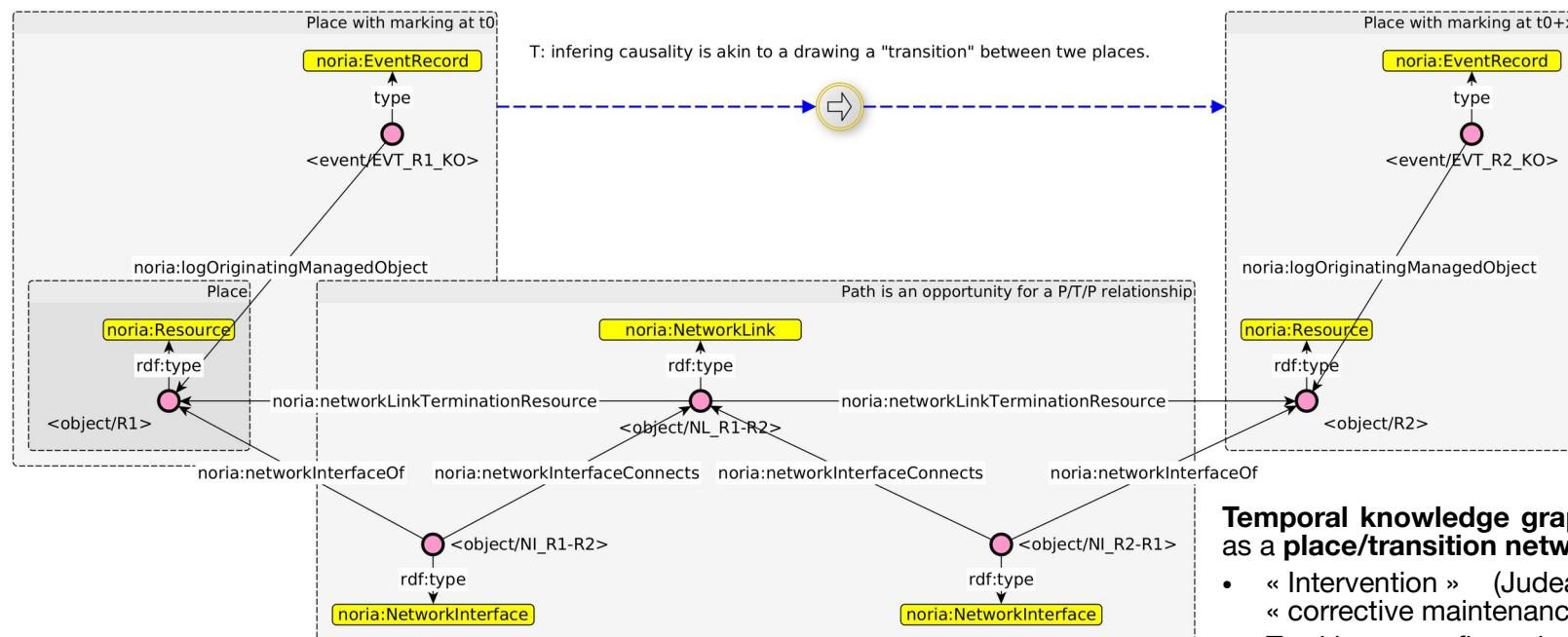
- Building the graph with **all incoming data**.
- Building the graph with **summarized data**, and ensure **uniqueness of object identifiers** across data stores.



# Causal Graphs & Knowledge Graphs

(General case) **Discovering causal graphs** from samples derived from a causal model: need for independence tests between variables (require a large amount of data to be accurate).

(NORIA case) **Not a « blind discovery »**: we already have some edges in the graph (even if they are not directed) + we also have access to temporal information, which is highly useful in causality (causes precede effects).

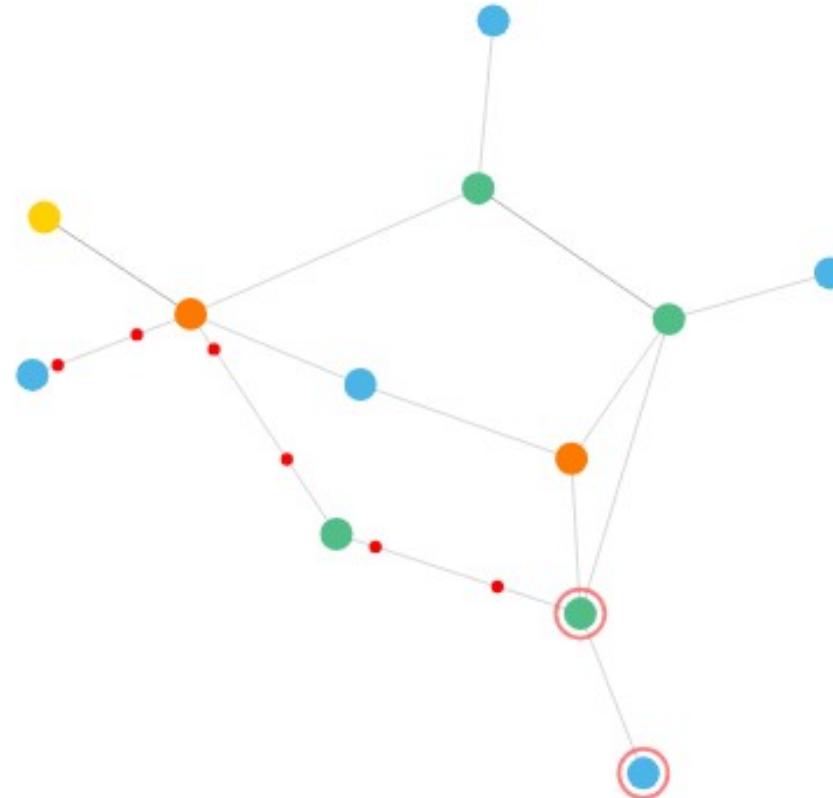


**Temporal knowledge graph** can be seen as a **place/transition network (PTnet)**

- « Intervention » (Judea Pearl) ⇔ « corrective maintenance action »
- Tracking reconfiguration actions on the network, it is possible to observe the dependency relationships between the states of network entities through the graph representation of the network.

|

# Graphical Root Cause Analysis

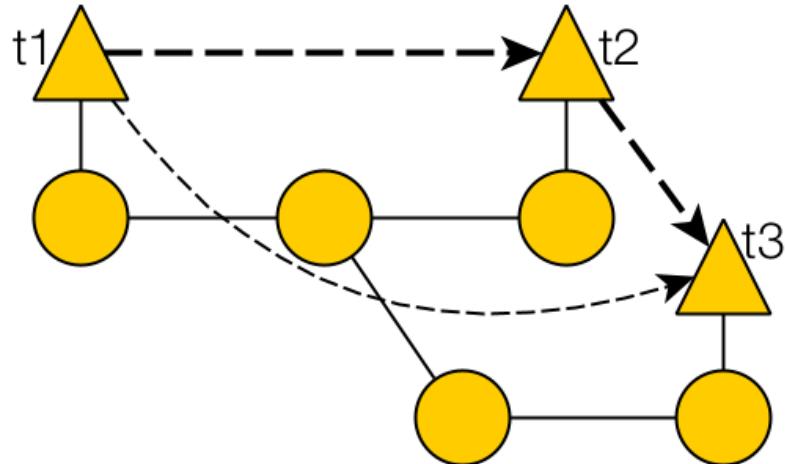


A prototype of the graphical root cause analysis view obtained by **projecting the procedural model** from the process mining step **onto the entities in the NORIA UI notebook**. The circled nodes highlight the *noria:Resource* and the *noria:EventRecord* likely responsible for the incident. The dotted lines emphasize the temporal sequence.

# Time-Ordered Contact Map

Without prior knowledge of event sequences: **disambiguating events** for which the occurrence time is close or identical.

We assume that the mechanism of **fault propagation** on the network is a **function of the distance** to be traveled in terms of the number of **network hops**.



$$\begin{pmatrix} 0_{1 \rightarrow 1} & \mathbf{2}_{1 \rightarrow 2} & 3_{1 \rightarrow 3} \\ 2_{2 \rightarrow 1} & 0_{2 \rightarrow 2} & \mathbf{3}_{2 \rightarrow 3} \\ 3_{3 \rightarrow 1} & 3_{3 \rightarrow 2} & 0_{3 \rightarrow 3} \end{pmatrix}$$

A toy example of a network topology with three events (triangular shapes with  $t_1 \leq t_2 \leq t_3$ ). The heavy dashed arcs represent « followed by » relationships (bold numbers in eq.) The light dashed arc represents the **transitive cause-effect relationship** of the  $t_1$  event to the  $t_2$  event, based on the composition  $(t_2 \rightarrow t_3) \circ (t_1 \rightarrow t_2)$ .

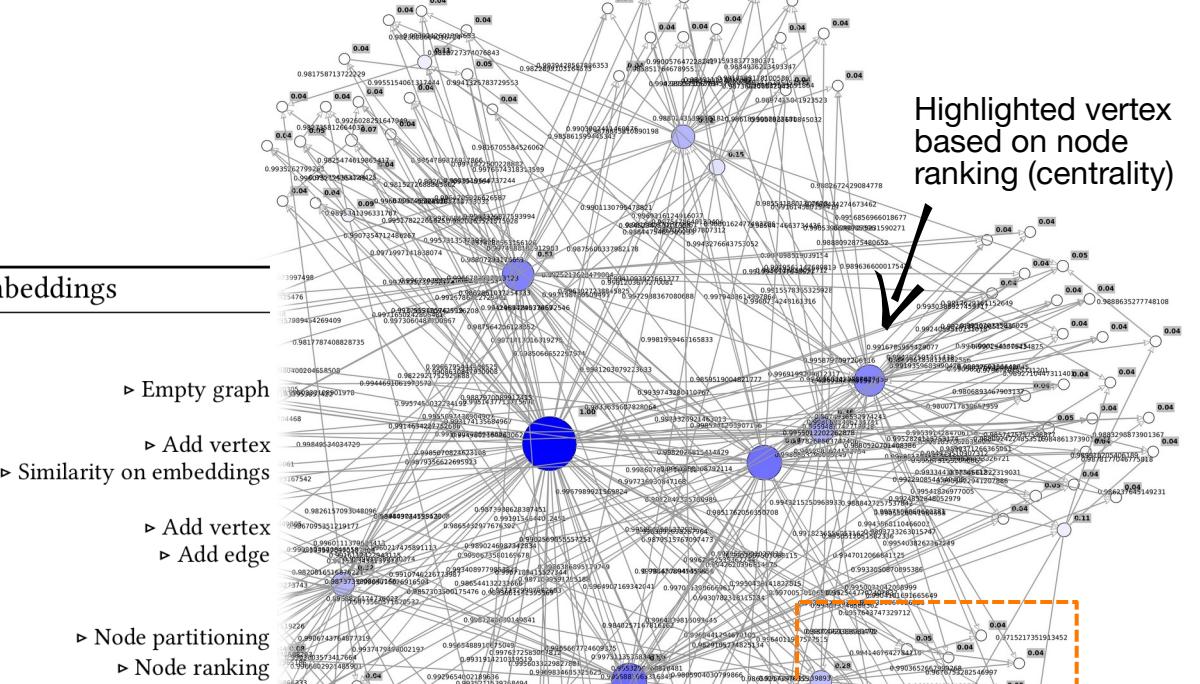
# Similarity Graph from Embeddings

## Algorithm 1 Similarity graph of entities embeddings

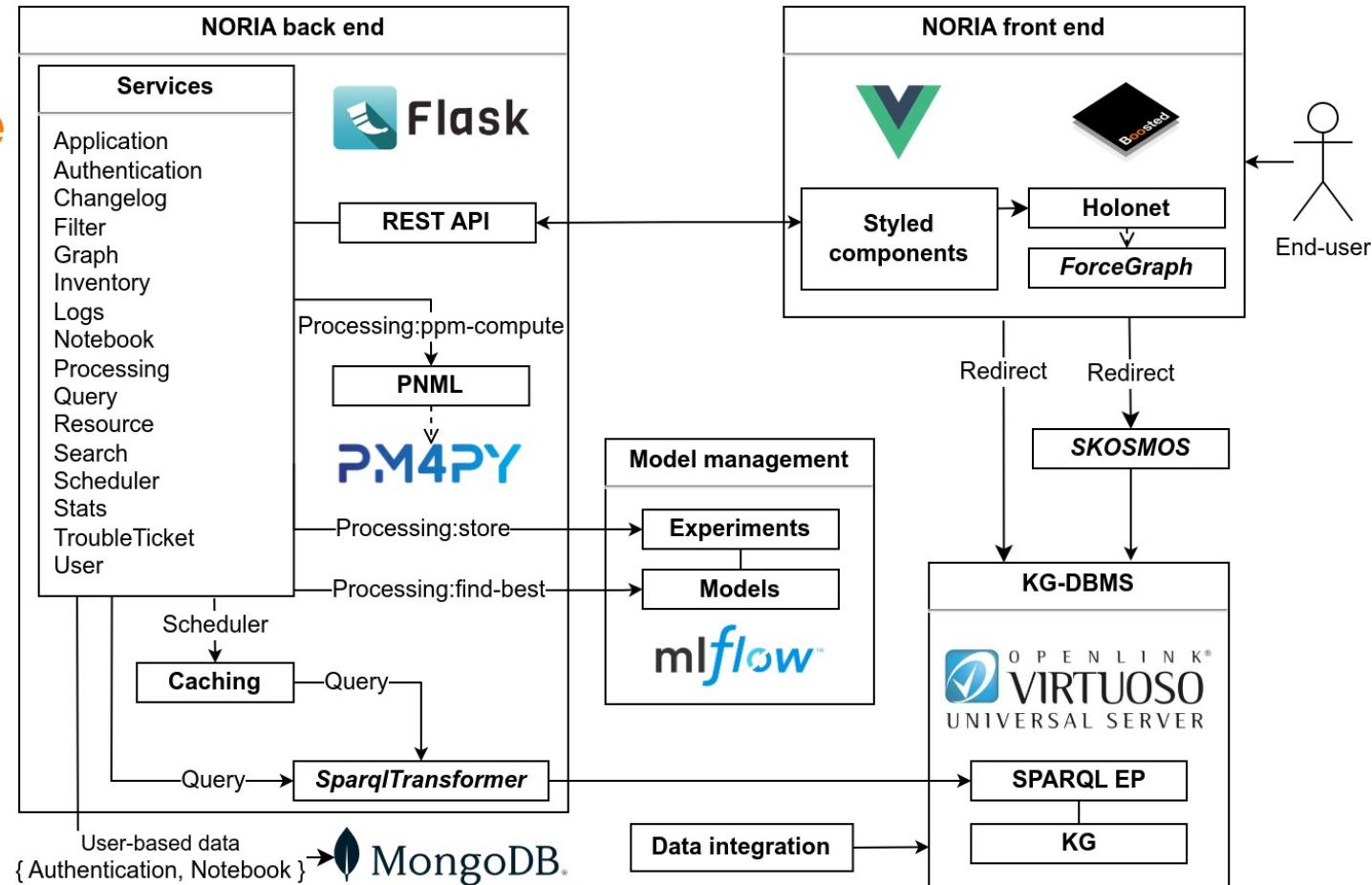
```

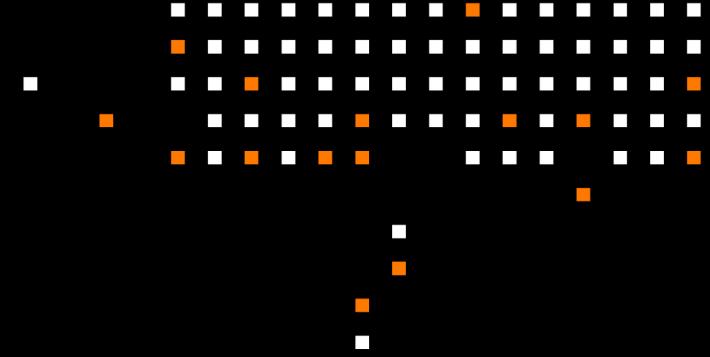
 $E \leftarrow$  embeddings entities
 $k \leftarrow$  number of entities for similarity
 $SG \leftarrow \emptyset$ 
for all  $e \in E$  do
     $SG \leftarrow e$ 
     $SIM \leftarrow \text{MostSimilar}_{\text{cosine}}(e, E, k)$ 
    for all  $e_{sim} \in SIM$  do
         $SG \leftarrow e_{sim}$ 
         $SG \leftarrow (e, e_{sim})$ 
    end for
end for
 $SG \leftarrow P_{\text{Louvain modularity}}(SG)$ 
 $SG \leftarrow R_{\text{Centrality}}(SG)$ 

```



# NORIA UI Architecture





# Thanks !