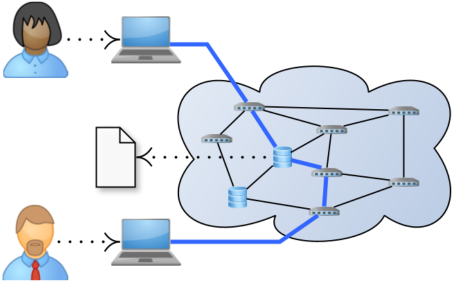
Designing NORIA: a Knowledge Graph-based Platform for Anomaly Detection and Incident Management in ICT Systems

KGCW @ ESWC 2023

Lionel Tailhardat, Orange, lionel.tailhardat@orange.com
Yoan Chabot, Orange, yoan.chabot@orange.com
Raphaël Troncy, EURECOM, raphael.troncy@eurecom.fr

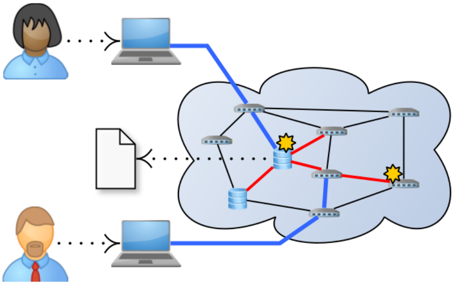
2023-05-28

Context & motivations: alarm spreading & heterogeneous networks



Scenario	Networking / online collaboration	Diagnosis	Situation understanding through causal models
Situation	Impaired network service	Real world	Alarm spreading phenomenon, heterogeneous networks (multi-technology, multi-vendor)
Observables	Alarms and logs		

Context & motivations: alarm spreading & heterogeneous networks



Scenario Networking / online collaboration

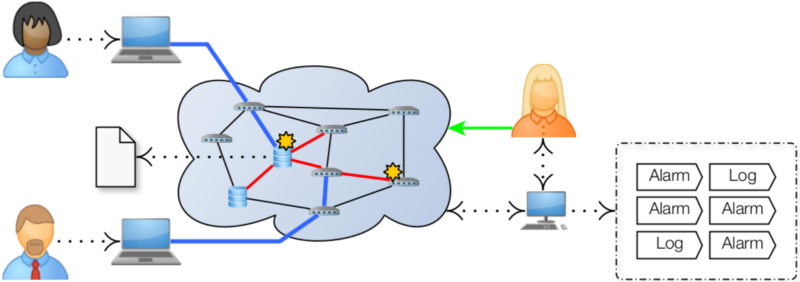
Situation Impaired network service

Observables Alarms and logs

Diagnosis Situation understanding through causal models

Real world Alarm spreading phenomenon, heterogeneous networks (multi-technology, multi-vendor)

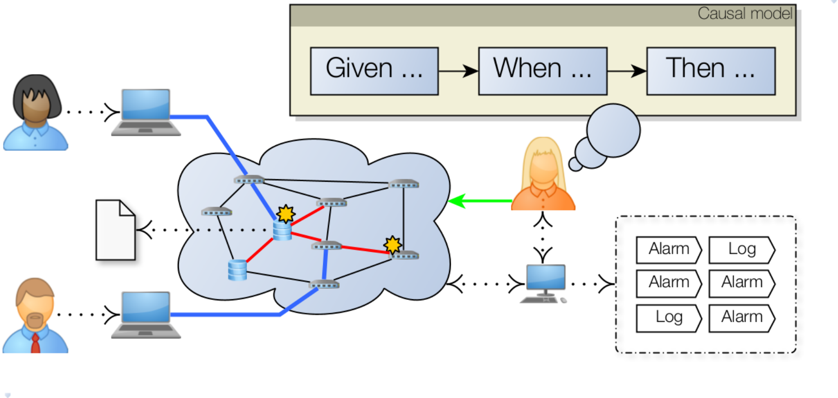
Context & motivations: alarm spreading & heterogeneous networks



Scenario	Networking / online collaboration
Situation	Impaired network service
Observables	Alarms and logs

Diagnosis	Situation understanding through causal models
Real world	Alarm spreading phenomenon, heterogeneous networks (multi-technology, multi-vender)

Context & motivations: alarm spreading & heterogeneous networks



Scenario Networking / online collaboration

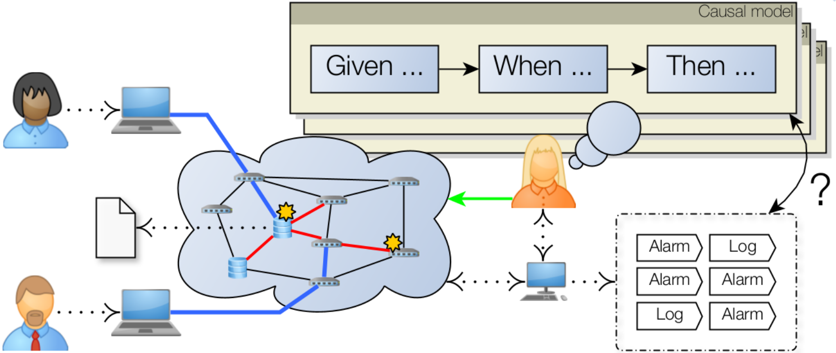
Situation Impaired network service

Observables Alarms and logs

Diagnosis Situation understanding through causal models

Real world Alarm spreading phenomenon, heterogeneous networks (multi-technology, multi-vender)

Context & motivations: alarm spreading & heterogeneous networks



Scenario	Networking / online collaboration
Situation	Impaired network service
Observables	Alarms and logs

Diagnosis	Situation understanding through causal models
Real world	Alarm spreading phenomenon, heterogeneous networks (multi-technology, multi-vendor)

Problem statement: contextualize network events and states efficiently?

Knowledge Structural (servers, routers, links), Functional (services, platforms), Dynamics (alarms, trouble tickets), Procedural (activity models).

3V Data Various sources, different formats (tabular, tree, graph, stream) and refresh periods (real-time $\rightarrow$ weekly).

Hypothesis Cross-referencing semantic representations from multiple sources enhances incident understanding.

Contributions

1. Design & implement a generic Knowledge Graph Construction tool chain, reusing/adapting well-known IT and SemWeb frameworks,
2. Evaluate the performance of the design, as well as the business value.

Problem statement: contextualize network events and states efficiently?

Knowledge Structural (servers, routers, links), Functional (services, platforms), Dynamics (alarms, trouble tickets), Procedural (activity models).

3V Data Various sources, different formats (tabular, tree, graph, stream) and refresh periods (real-time $\rightarrow$ weekly).

Hypothesis Cross-referencing semantic representations from multiple sources enhances incident understanding.

Contributions

1. Design & implement a generic Knowledge Graph Construction tool chain, reusing/adapting well-known IT and SemWeb frameworks,
2. Evaluate the performance of the design, as well as the business value.

Problem statement: contextualize network events and states efficiently?

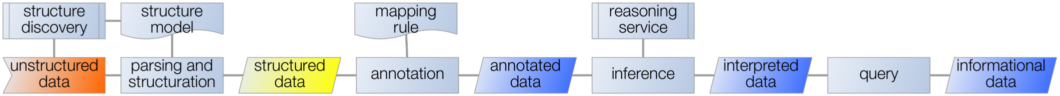
Knowledge Structural (servers, routers, links), Functional (services, platforms), Dynamics (alarms, trouble tickets), Procedural (activity models).

3V Data Various sources, different formats (tabular, tree, graph, stream) and refresh periods (real-time $\rightarrow$ weekly).

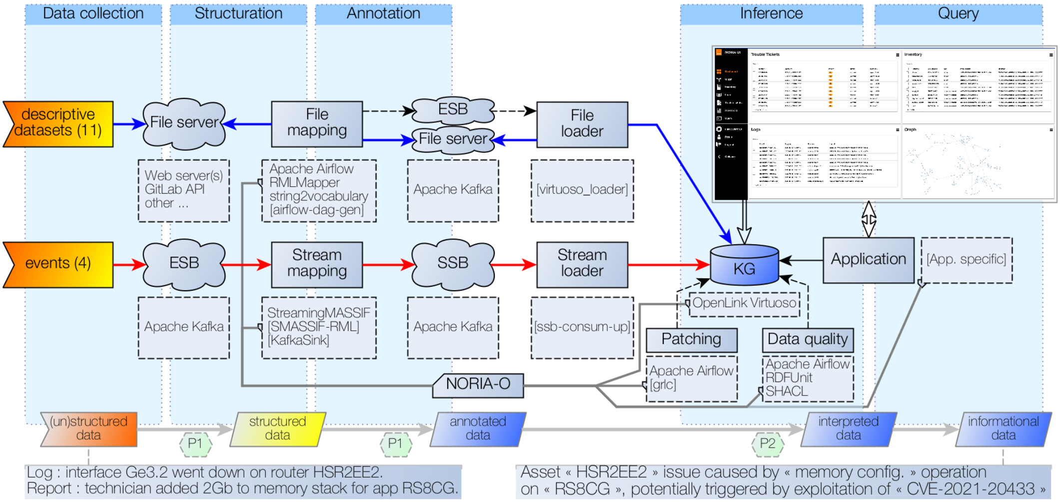
Hypothesis Cross-referencing semantic representations from multiple sources enhances incident understanding.

Contributions

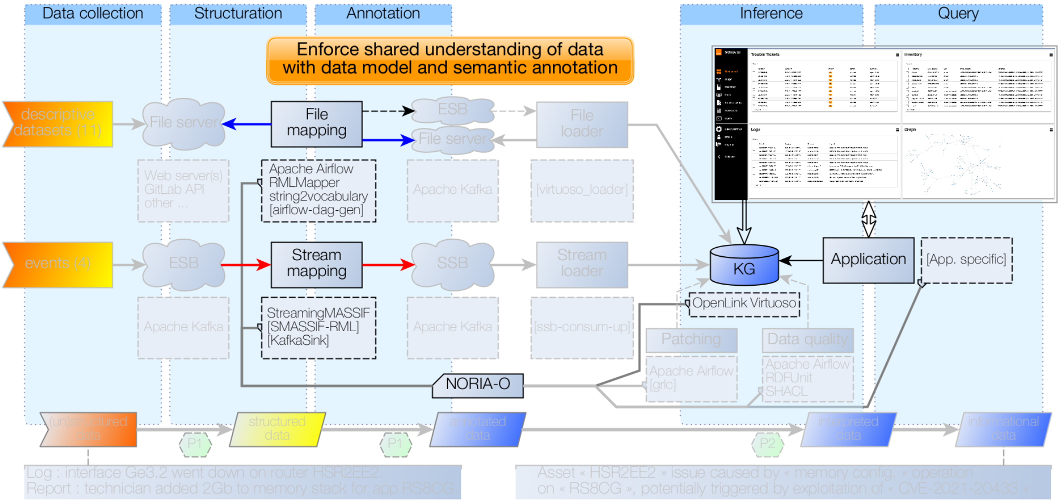
1. Design & implement a generic Knowledge Graph Construction tool chain, reusing/adapting well-known IT and SemWeb frameworks,
2. Evaluate the performance of the design, as well as the business value.



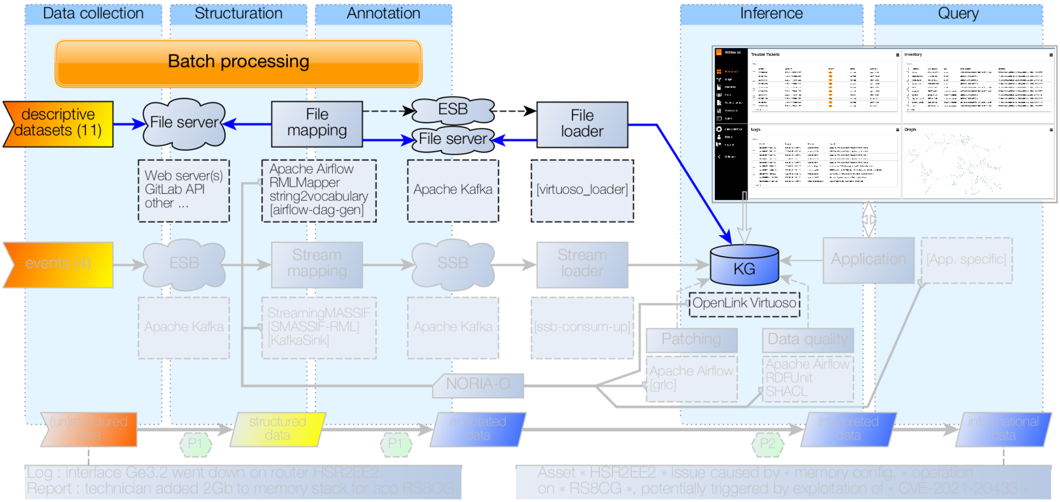
Contribution: the NORIA data integration architecture



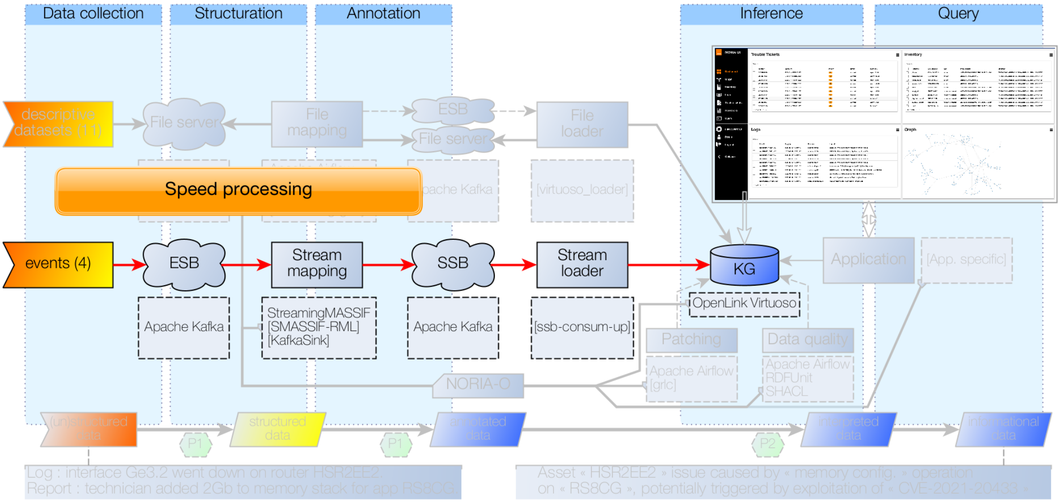
Contribution: the NORIA data integration architecture



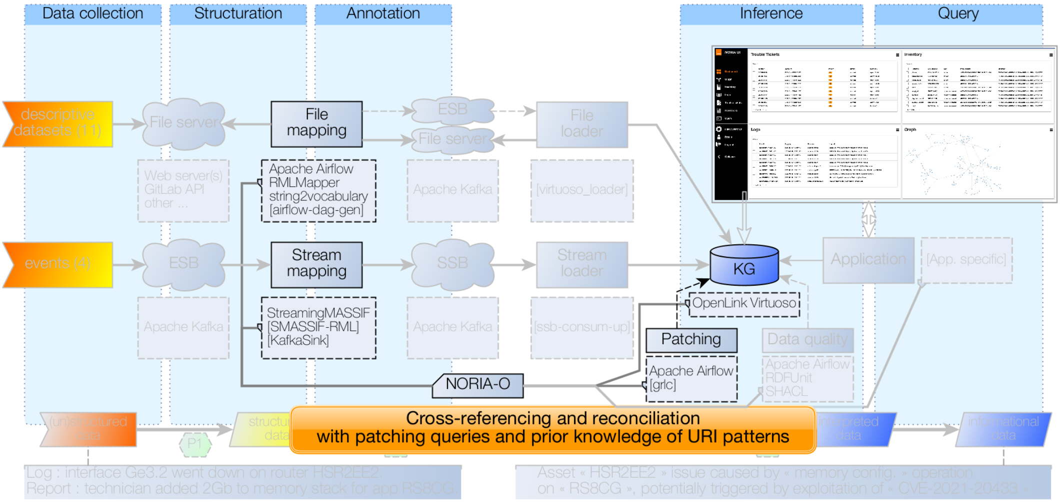
Contribution: the NORIA data integration architecture



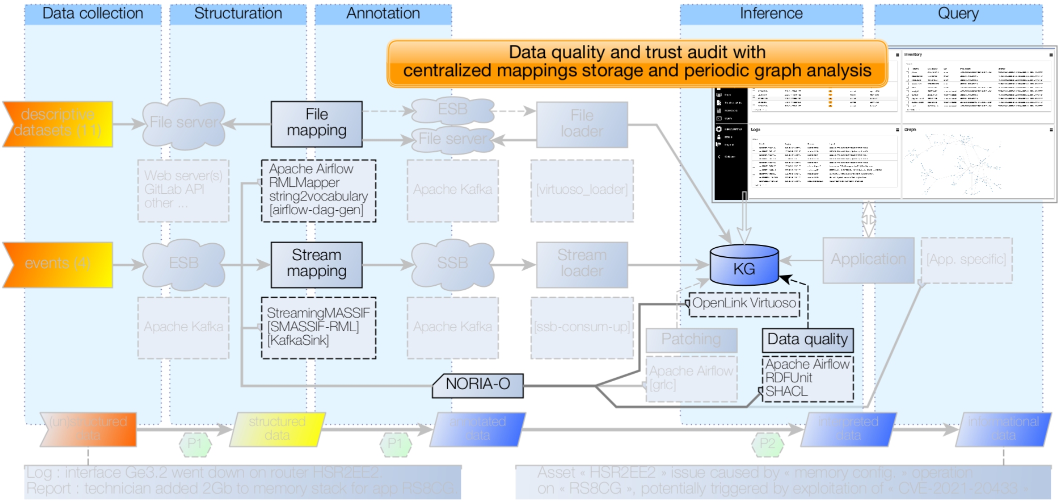
Contribution: the NORIA data integration architecture



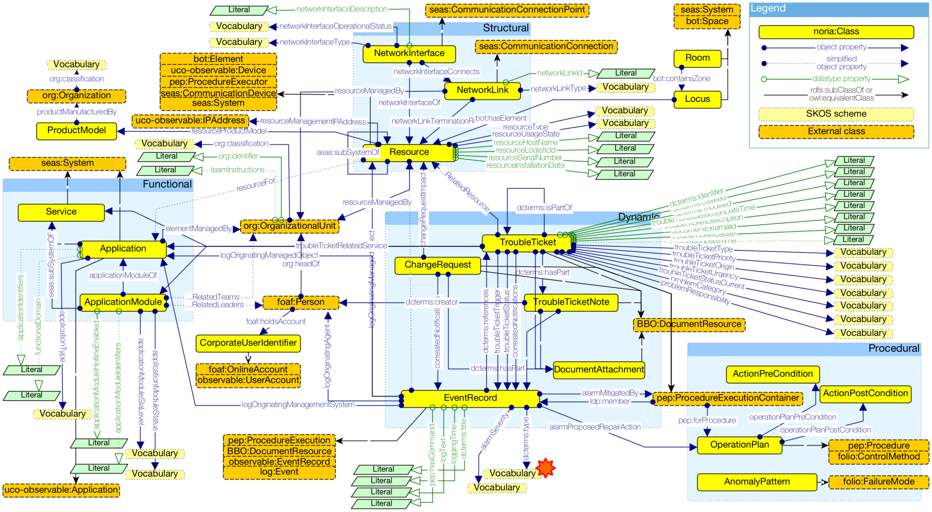
Contribution: the NORIA data integration architecture



Contribution: the NORIA data integration architecture

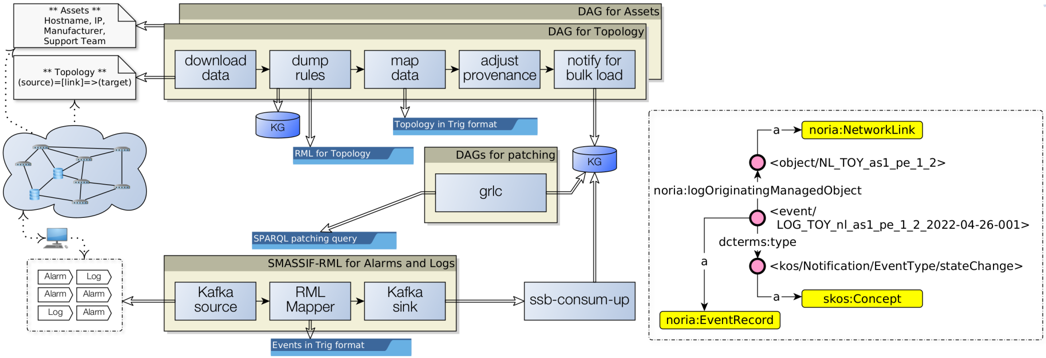


Overview of the NORIA-O v0.2 data model

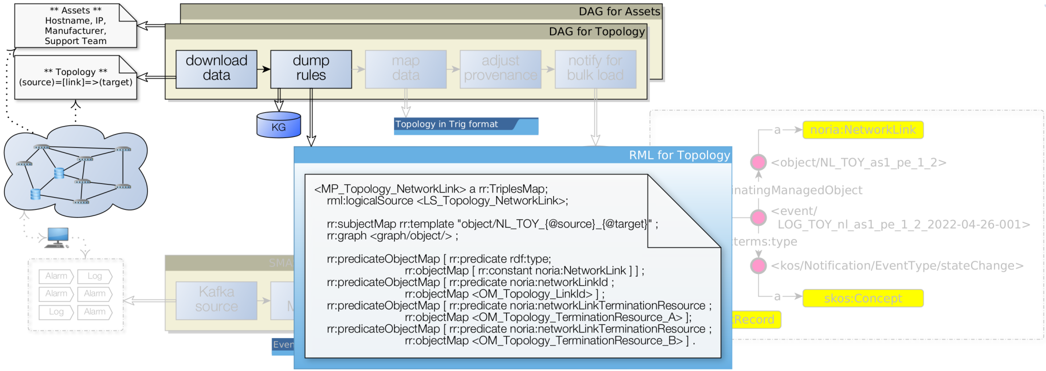


NORIA-O: https://w3id.org/noria/ (open source release under BSD-4 license) [SWJ 3334-4548]

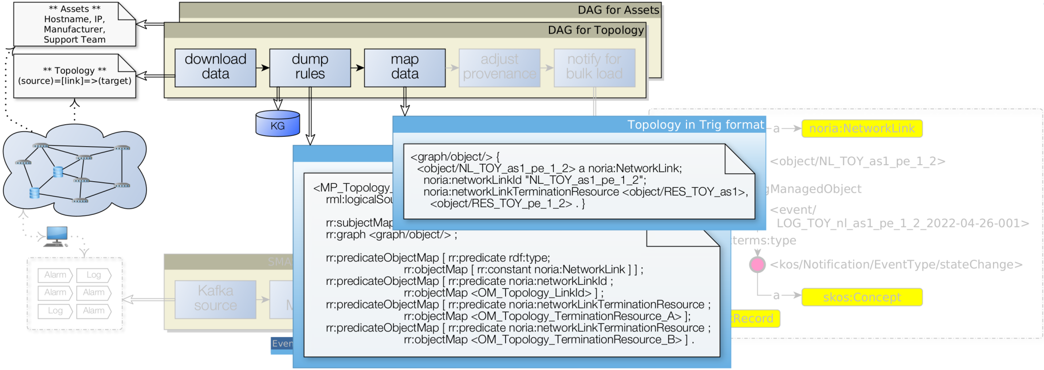
Knowledge graph construction example



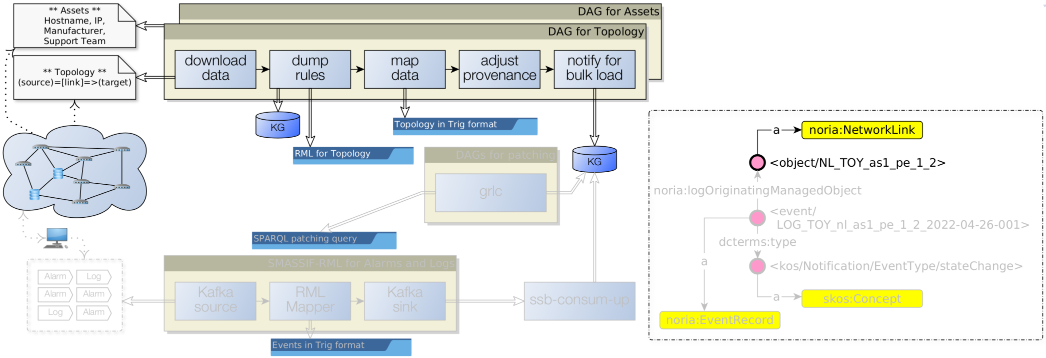
Knowledge graph construction example



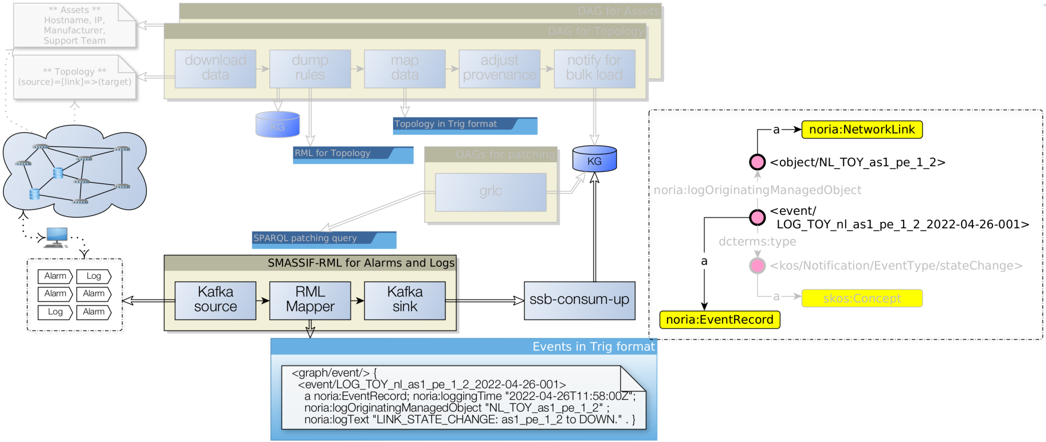
Knowledge graph construction example



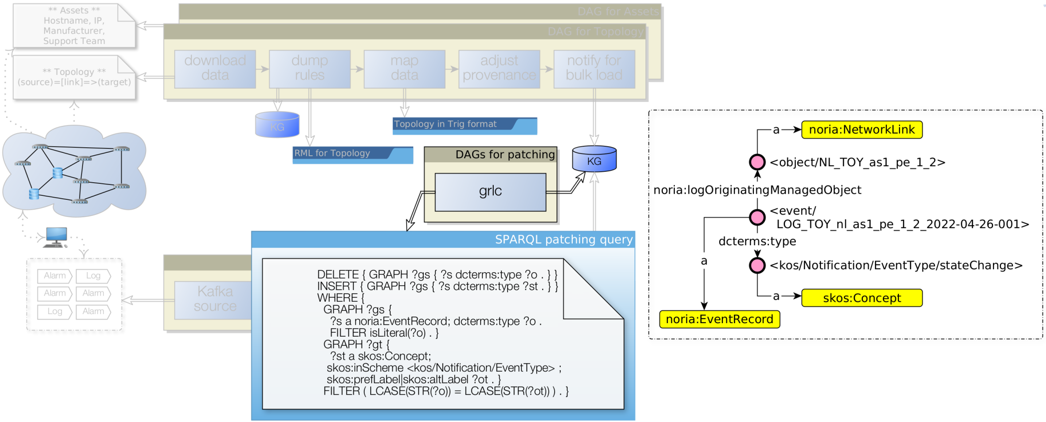
Knowledge graph construction example



Knowledge graph construction example



Knowledge graph construction example



Performance

Data integration

- 15 sources $\xrightarrow{39 \text{ rr:TriplesMap}}$ 4M triples (400K entities), including streamed events spanning over 111 days.
  - Batch processing: performance $\sim$ “map data” (w/o join) and “adjust provenance” stages,
  - Stream processing: effective, load testing is needed to go further.
- 42 patching SPARQL queries: 16 literal2SKOS , 19 literal2URI, 7 addShortcut.

	AAA security groups (small)		Users (medium)		Equipment database (big)		Unit
Input data size	0.16		2.4		45.5		[Mb]
Download data	0.44	6.63 %	0.95	1.54 %	3.32	0.69 %	[s]
Dump rules	0.14	2.11 %	0.19	0.31 %	0.15	0.03 %	[s]
Preprocessing	0.19	2.86 %	9.46	15.37 %	8.66	10.83 %	[s]
Map data	3.27	49.25 %	8.54	13.87 %	79.97	16.70 %	[s]
Adjust provenance	2.27	34.19 %	40.66	66.05 %	374.26	78.16 %	[s]
Notify for loading	0.27	4.07 %	0.29	0.47 %	0.29	0.06 %	[s]
Data bulk load	0.05	0.75 %	1.46	2.37 %	12.17	2.54 %	[s]
Prov. bulk load	0.01	0.15 %	0.01	0.02 %	0.02	0.00 %	[s]
Total time	6.64		61.56		478.84		[s]
Output data	0.52		21		222		[Mb]
	5 110		244 532		2 415 676		[Triples]
Throughput	769.58		3 972.25		5 044.85		[Triples/s]

Summary & future work

Problem	Integrating and linking heterogeneous data to facilitate the diagnosis and management of network incidents.
Our approach	Lambda architecture using SemWeb technologies, centralized mappings storage, patching and reconciliation tasks.
Next	Kappa architecture, ETL process as a graph, anomaly detection, cooperative decision making.

Paper

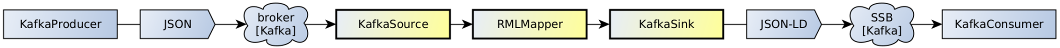
Designing NORIA: a Knowledge Graph-based Platform for Anomaly Detection and Incident Management in ICT Systems.
https://w3id.org/kg-construct/workshop/2023/resources/paper3.pdf

Code repository

- SMASSIF-RML
  https://github.com/Orange-OpenSource/SMASSIF-RML
- ssb-consum-up
  https://github.com/Orange-OpenSource/ssb-consum-up
- grlc
  https://github.com/Orange-OpenSource/grlc
- NORIA-O
  https://w3id.org/noria/

Appendices

Where do I start? The SMASSIF-RML quick start



1. Git clone the project to your computer

```
git clone https://github.com/Orange-OpenSource/SMASSIF-RML.git
cd SMASSIF-RML
```

2. Install and build the SMASSIF-RML tool set, then start the demo pipeline

```
make install-dependencies
mvn package
make start-kafka
make demo-dsm
```

3. Observe mapping in CLI output

Where do I start? The ssb-consum-up quick start

1. Git clone the project to your computer

```
git clone https://github.com/Orange-OpenSource/ssb-consum-up.git
cd ssb-consum-up
```

2. Install and the ssb-consum-up tool set, then start the demo pipeline

```
make install-dev-tools
make start-kafka
make start-virtdb
make start-scu-script
make start-producer
```

3. Observe the ssb-consum-up logs for data consume/update notifications
4. Get the inserted demo data from the graph store

```
make get-demo-data
```

Evaluating NORIA-O with authoring tests

Evaluation set 26 Competency Questions (CQs), available at https://w3id.org/noria/cqs/, translated into 25 Authoring Tests (SPARQL queries).

Evaluation results three different situations summarized as “OK” (16/26), “AI” (9/26) and “Extension” (1/26).

Evaluation results	#CQs	Remarks
OK	16/26	Answered using a single or several simple SPARQL queries and the ontology.
AI	9/26	Require the implementation of more complex AI-based algorithms such as anomaly detection algorithms.
Extension	1/26	Require the introduction of new concepts or relations via an extension of the NORIA-O model.

“OK” example “Which entity (resource/application/site) is concerned by a given incident?”

“AI” example 1 “What was the root cause of the incident?”,
→ 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.

“AI” example 2 “What are the vulnerabilities and the associated risk levels of this infrastructure?”,
→ 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.

“Extension” example “What is the financial cost of this incident if it occurs?”,
→ involves information about the cost of an incident.

Who's who

Lionel Tailhardat AI R&D Engineer

- Dynamic Systems, Dependability and Knowledge Engineering
- genears.github.io

Dr. Raphaël Troncy Associate Professor

- Knowledge Engineering, Knowledge Graphs and Data Science
- www.eurecom.fr/~troncy

Dr. Yoan Chabot AI Researcher

- Knowledge Engineering, Knowledge Graphs and Data Science
- yoanchabot.github.io

Orange

Intl. Telecommunication infrastructure and service provider (and more ...)

- www.orange.com
- hellofuture.orange.com

EURECOM

Graduate School and Research Center in Digital Science

- www.eurecom.fr

Our proposition: combine AI and Knowledge Engineering techniques for Complex Networks Resilience and Data Security concerns.