

# Semantic Metadata **Annotation** for Network **Anomaly** Detection

draft-netana-nmop-network-anomaly-semantics-01

Helps to test, validate and compare outlier detection, supports supervised and semi-supervised machine learning development, enables data exchange among network operators, vendors and academia, and make anomalies for humans apprehensible

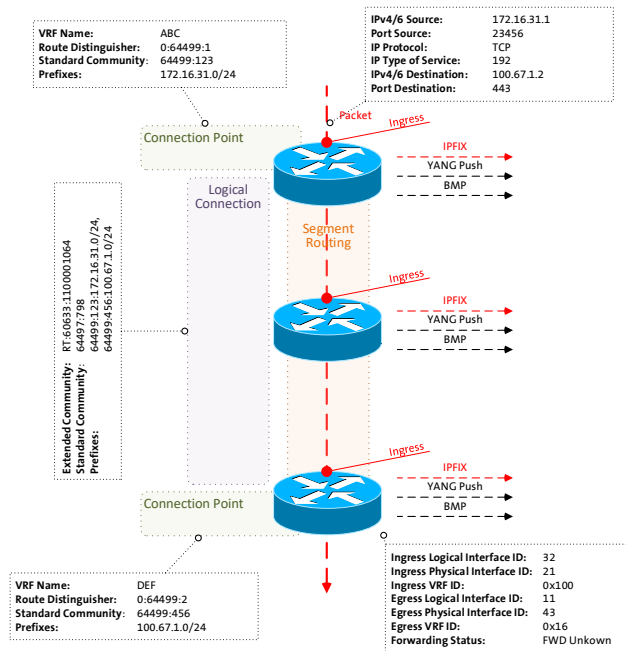
thomas.graf@swisscom.com  
wanting.du@swisscom.com  
alex.huang-feng@insa-lyon.fr  
vincenzo.riccobene@huawei-partners.com  
antonio.roberto@huawei.com

05. March 2024

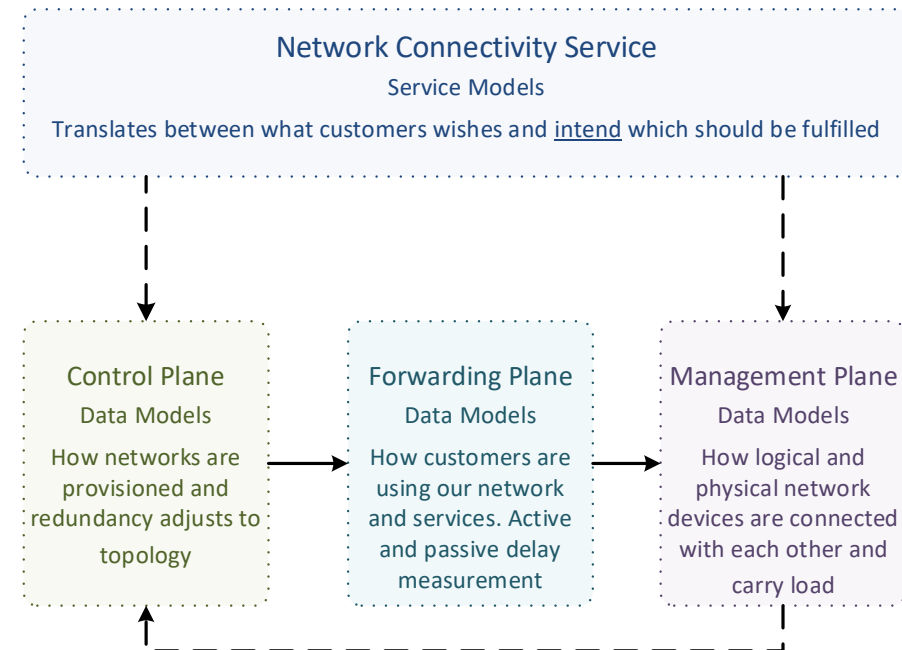
# What to monitor

Which operational metrics are collected

« Network operators **connect customers in** routing tables called **VPN's** »

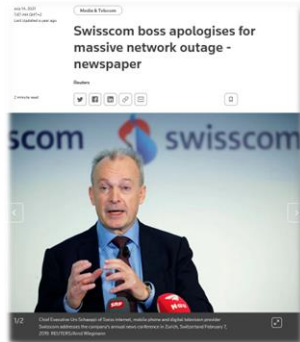
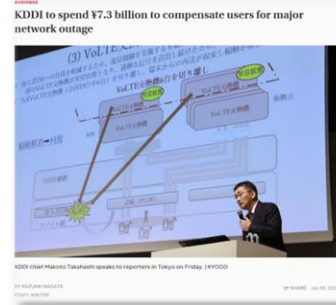


« Network Telemetry (RFC 9232) describes how to collect data from **all 3 network planes** efficiently »



# Why to automate monitoring

Recognize network incidents faster than humans can



05 FEB 2023 | 08:23 AM UTC

Italy: TIM internet services interruption reported nationwide Feb. 5

TIM internet services interruption reported in Italy Feb. 5. Likely communication disruptions.

Informational Communications/technology Transportation ITA



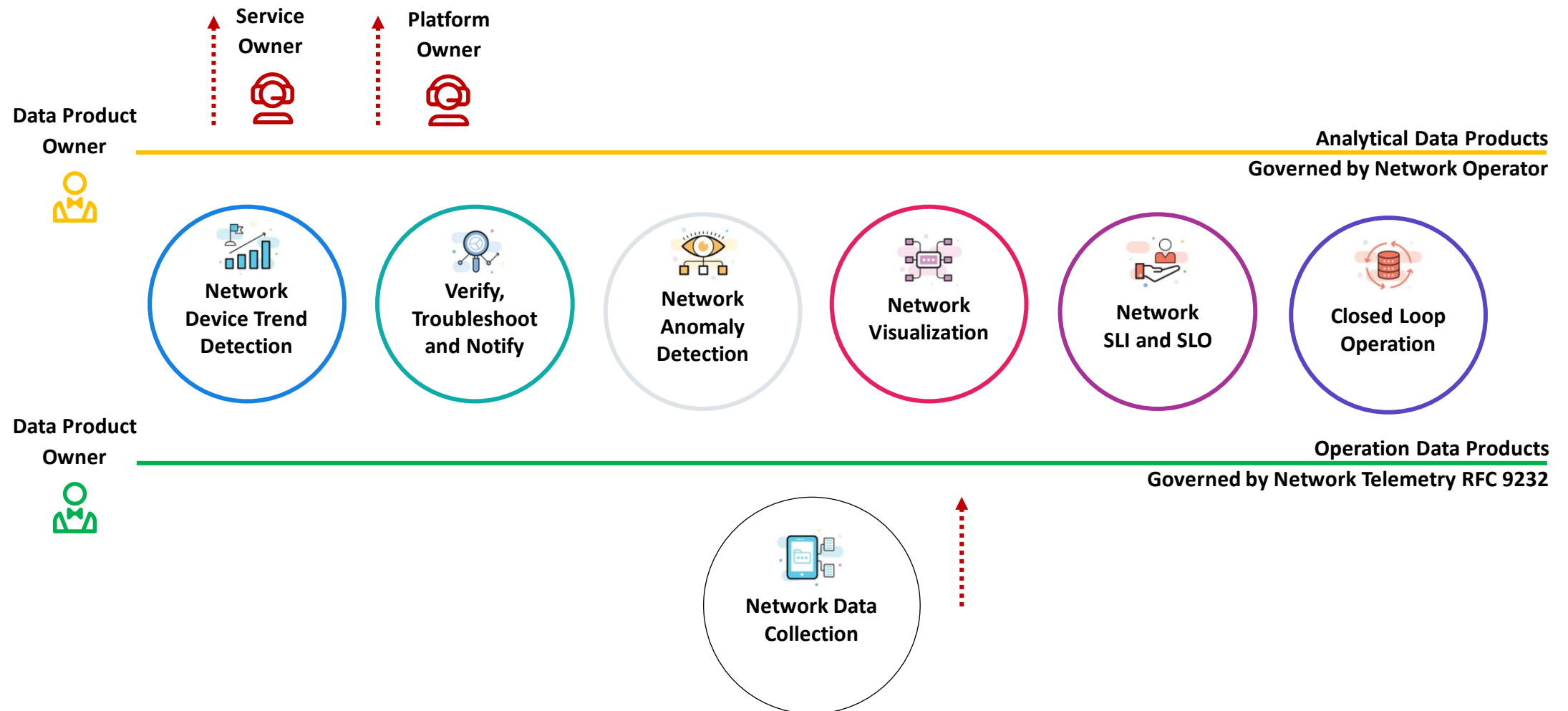
Facebook outage: what went wrong and why did it take so long to fix after social platform went down?



« Customers are **always connected, when VPN's changing**, regardless due to operational or configurational reasons, network operators are **late to react** due to **missing visibility and automation** »

# How to organize and collaborate with data

The Data Mesh Architecture enables Network Analytics use



# What does Network Anomaly Detection mean

Monitor changes



## Network Anomaly Detection

**For VPNs**, Network Anomaly Detection **constantly monitors and detects any network or device topology changes**, along with their associated forwarding consequences for customers as outliers. Notifications are sent to the Network Operation Center before the customer is aware of service disruptions. **It offers operational metrics for in-depth analysis**, allowing to understand on which platform the problem originates and facilitates problem resolution.



### Answers

What changed and when, on which connectivity service, and how does it impact the customers?



### Focuses

Provides meaningful connectivity service impact information before customer is aware of and support in root-cause analysis.



### Data Mesh

Consumes operational real-time Forwarding Plane, Control Plane and Management Plane metrics and produces analytical alerts.



### Direction

From connectivity service to network platform.

« A more detailing paper  
will be submitted soon to  
IEEE Transactions on  
Network and Service  
Management»



# What our motivation is

Automate learn and improve

From network incidents postmortems we network operators **learn and improve** so does network anomaly detection and supervised and semi-supervised machine learning.

The more network incidents are observed, the more we can improve. With more incidents the **postmortem process needs be automated, let's get organized** first by defining human and machine-readable metadata semantics and annotate operational and analytical data.

Let's get further organized by exchanging standardized labeled network incident data among network operators, vendors and academia to **collaborate on academic research**.

« The community working on Network Anomaly Detection is probably the only group **wishing for more network incidents** »

# What is a symptom and how to categorize them

From action to reason to cause

**Action:** Which action the network node performed for a packet in the forwarding plane, a path or adjacency in the control plane or state or statistical changes in the management plane.

**Reason:** For each reason one or more actions describing why this action was used. From drop unreachable, administered, and corrupt in forwarding plane, to reachability withdraw and adjacency teared down in control plane, to Interface down, errors or discard in management plane.

**Cause:** For each reason one or more causes describes why the action was chosen. From missing next-hop and link-layer information in forwarding plane, to reachability withdrawn due to peer down or path no longer redistributed.

« Symptoms are categorized in **which plane** they have been **observed**, their **action, reason and cause** »



# Questions to the audience

Do you care?

**Network Operators:** Do you agree that today's actions; traffic is dropped, path is withdrawn and interface down, are always exposed through Network Telemetry. But reasons and causes, dropped due to unreachable next-hop, withdrawn due to peer down, interface down due to missing signal, are rarely exposed to telemetry would be most interesting?

**Network Vendors:** Is the assumption correct that a when network service process, routing process and withdrawing a path occur, most of the time the vendor knows why it acts that way, and could potential make this reason and cause information available?

**Academia:** Would it help if network operators would provide well defined labeled operational and analytical data to enable and validate their research?

**Everybody:** Should these symptoms be clearly described and standardized for a common terminology so that operators, researchers and anomaly detection systems alike understand their meaning and learn and act accordingly?

# Outliers in Anomaly Detection

From global to contextual to collective

**Global outliers:** An outlier is considered "global" if its behavior is outside the entirety of the considered data set.

**Contextual outliers:** An outlier is considered "contextual" if its behavior is within a normal (expected) range, but it would not be expected based on some context. Context can be defined as a function of multiple parameters, such as time, location, etc.

**Collective outliers:** An outlier is considered "collective" if the behavior of each single data point that are part of the anomaly are within expected ranges (so they are not anomalous, it's either a contextual or a global sense), but the group taking all the data points together, is.

« **Collective outliers** are important because networks are connected. Through **different planes interconnected** symptoms from various angles can be observed »

# Annotate Operation Data

## YANG Module

```
module: ietf-symptom-semantic-metadata
```

```
+--rw symptom
  +--rw id          yang:uuid
  +--rw event-id    yang:uuid
  +--rw description  string
  +--rw start-timeyang:date-and-time
  +--rw end-time    yang:date-and-time
  +--rw confidence-score float
  +--rw concern-score? float
```

```
+--rw tags* [key]
  | +--rw key    string
  | +--rw value  string
```

```
+--rw (pattern)?
  | +--:(drop)
  | | +--rw dropempty
  | +--:(spike)
  | | +--rw spike          empty
  | +--:(mean-shift)
  | | +--rw mean-shift     empty
  | +--:(seasonality-shift)
  | | +--rw seasonality-shift empty
  | +--:(trend)
  | | +--rw trend          empty
  | +--:(other)
  | +--rw other            string
```

```
+--rw source
  +--rw (source-type)
  | +--:(human)
  | | +--rw human    empty
  | +--:(algorithm)
  | | +--rw algorithm empty
  +--rw name?        string
```

- **Symptoms** describe what changed in the network for what reason and cause with which concern score from when to when.
- **Tags** describes in which network plane, which action, reason and cause was observed.
- **Pattern** describes the measurement pattern over time of the time series data.
- **Source** describes which system **observed** the outlier. A human or a network anomaly detection system.

# Network Anomaly Postmortem Lifecycle

Relationship to draft-netana-nmop-network-anomaly-lifecycle

**draft-netana-nmop-network-anomaly-lifecycle** defines how the symptoms and pattern applied in the Network Anomaly Postmortem Lifecycle

## 4. Lifecycle of a Network Anomaly

The lifecycle of a network anomaly can be articulated in three phases, structured as a loop: Detection, Validation, Refinement.

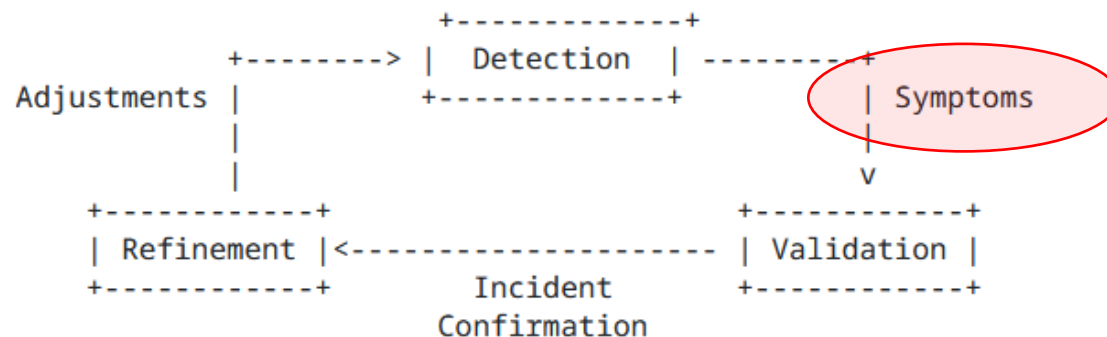


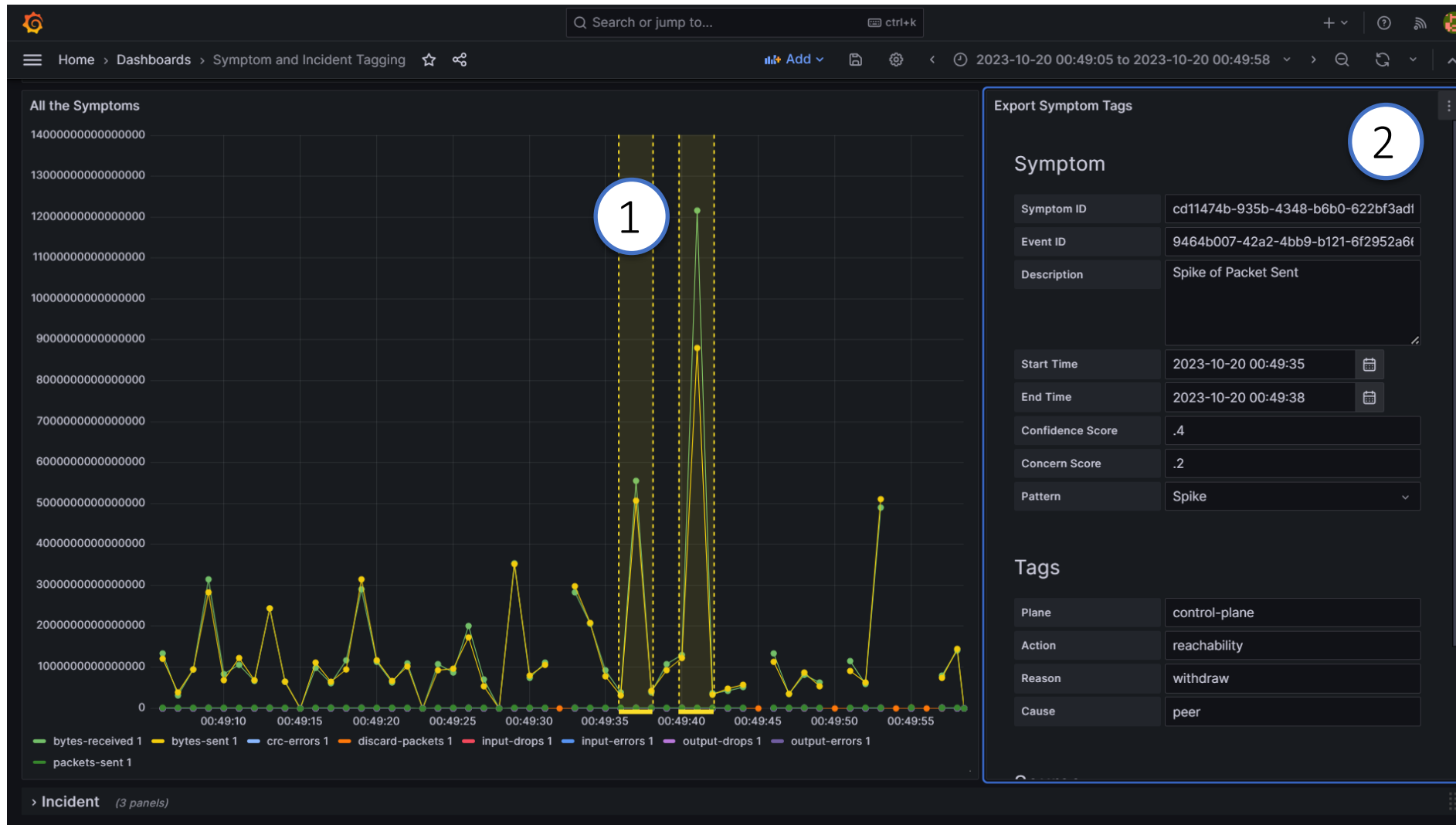
Figure 1: Anomaly Detection Refinement Lifecycle

Each of these phases can either be performed by a network expert or an algorithm or complementing each other.

« Network Anomaly Postmortems helps to identify shortcomings and refine the Anomaly Detection System **to reduce the number of false positives** »

# IETF 118 Hackathon – Antagonist

## Labelling a Symptom in Grafana



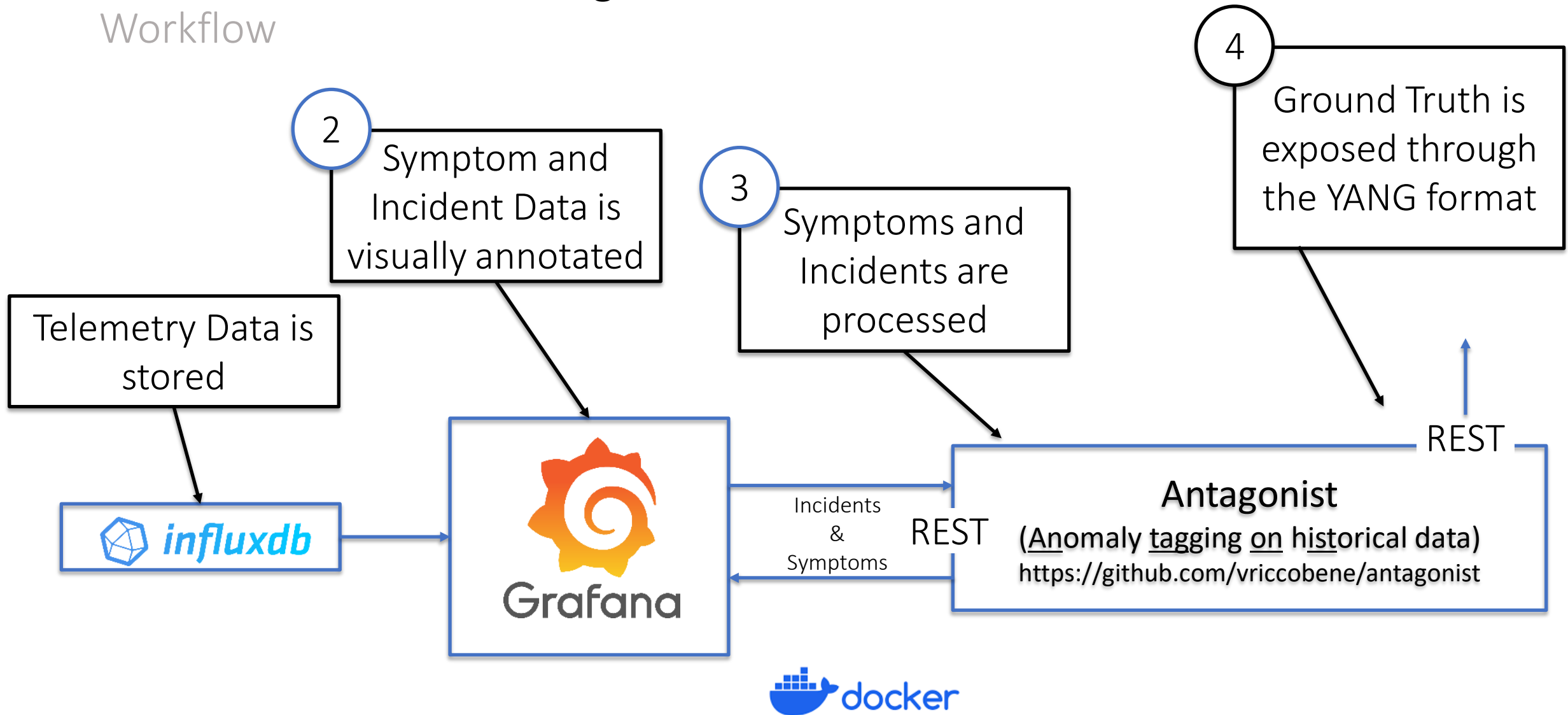
(1) Vertical dotted lines are the tagged symptoms.

(2) Once the symptom is selected, the user can add all the details.

Once the symptom is defined it gets submitted to Antagonist.

# IETF 118 Hackathon - Antagonist

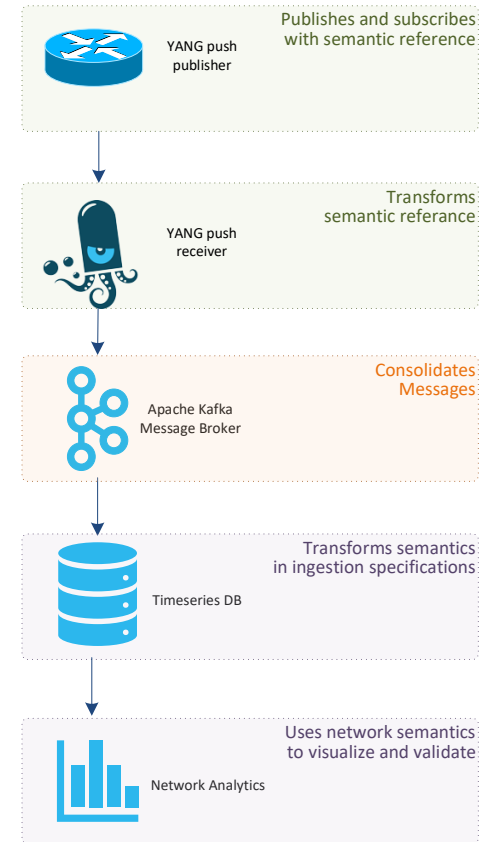
## Workflow



# Semantic Metadata Annotation for Network Anomaly Detection

## Next steps

- **This work relates to the Network Anomaly Detection topic in the NMOP charter.**
- It bridges network and data engineering, operator, vendors and academia, domains by having the **semantics and ontology of network symptoms for operational and analytical data defined.**
- This work will unveil what is missing in Network Telemetry data and provide input to other documents such as draft-davis-nmop-incident-terminology to enable a more detailed and holistic view for networks.
- **Do you realize the benefit of having standardized semantic metadata annotation for Network Anomaly Detection and how it helps network operators, vendor and academia to collaborate?**
- **-> What are your thoughts and comments?**
- **-> We request NMOP working group adoption.**



thomas.graf@swisscom.com  
wanting.du@swisscom.com  
alex.huang-feng@insa-lyon.fr  
vincenzo.riccobene@huawei-partners.com  
antonio.roberto@huawei.com

05. March 2024