

# A Data Manifest for Contextualized Telemetry Data

[draft-ietf-opsawg-collected-data-manifest](#)

B. Claise (Huawei), J. Quilbeuf (Huawei), D. Lopez  
(Telefonica), I. Dominguez (Telefonica), T. Graf  
(Swisscom)

IETF 119, OPSAWG

# Goal & Problem Statement

- Goal is not to expose new information via YANG but rather to define what needs to be kept as metadata (or Data Manifest) to ensure that the data can still be interpreted correctly even:
  - if the source device is not accessible (from the collection system)
  - If the source device has been updated or has a new configuration
- End goal: analyze the data, from the data collection system, with the proper context, for anomaly detection and, in the end, closed loop automation
- Per-node capability discovery exists
  - YANG Modules describing Capabilities for Systems and Datastore Update Notifications, [RFC9196](#) + YANG Instance Data File Format, [RFC9195](#)
  - Per-Node Capabilities for Optimum Operational Data Collection , [draft-claise-netconf-metadata-forcollection-03](#)
- But how were data actually metered, under which circumstances?

# Data Collection Vantage Point

No updates since t1

- Problem: Telemetry issue?
- Problem: Sender overloaded?
- Problem: bug?
- No problem: long telemetry cadence?
- No problem: “on-change”?
- No problem: “suppress-redundancy”?

Counter1 = 42  
Status = UP



End goal: analyze the data, from the data collection system, with the proper context, for anomaly detection and, in the end, closed loop automation

# Proposal: Data Manifest

- Data Manifest composed of 2 YANG models for storing the context:
  - **Platform Manifest**: part of the Data Manifest that completely characterizes the platform producing the data.
  - **Data Collection Manifest**: part of the Data Manifest that completely characterizes how and when the telemetry was metered.
- “MUST be streamed all with the data and stored along with the collected data.”
- “In case the data are moved to different place (typically a database), the data manifest MUST follow the collected data.”

# Changes

- Minor changes, clarify that module is network/Controller level
- Change InfluxDB example to TSDB example using notation from draft-kll-yang-label-tsdb

```
In our example, The link between a collected datapoint and the corresponding Platform Manifest is done via the common "device" tag. In order to link a datapoint with the corresponding Data Collection Manifest, the collector can add fields to specify where the Data Collection Manifest is located for that specific datapoint. For instance, the same datapoints as in Figure 4 could be stored as in Figure 6.
```

```
admin_status,device="PE1",interface="gig1" val=T,subId=4242
sent_bytes,device="PE1",interface="gig1" val=1234,subId=4243
```

Figure 6: Storing datapoints with data manifest

```
In our example, we can retrieve the Platform Manifest associated to a collected datapoint by looking for datapoints that have the label "platforms_platform_id" equal to the value of the host for that collected datapoint. In order to link a datapoint with the corresponding Data Collection Manifest, we need to add an additional label for the subscription id. For instance, the same datapoints as in Figure 4 could be stored as in Figure 6.
```

```
* Metric: interfaces_interface_enabled
* Value: True
* Labels:
  - host: "PE1"
  - interfaces_interface_name: "eth0"
  - data_collections_data_collection_yang_push_subscriptions_subscription_id: 4242
--
* Metric: interfaces_interface_statistics_in_octets
* Value: 1234
* Labels:
  - host: "PE1"
  - interfaces_interface_name: "eth0"
  - data_collections_data_collection_yang_push_subscriptions_subscription_id: 4243
```

Figure 6: Storing datapoints with information to retrieve the Data Manifest

# Next Steps

- Work on the YANG modeling issue
  - Current version in draft is based on copy-pasted modules)
  - draft-jouqui-yang-full-include (presentation in NETMOD on Thursday) would be a normative reference.
- Try to align with draft-lindblad-tlm-philatelist

# Open Questions

- How do we make the link with inventory efforts going on (IVY, RFC8345)? Notably for the platform-id, using the node id for RFC8345?
- Are we missing anything in the current form of the data manifest?

Feedback, suggestions, issues, PRs:

<https://github.com/JeanQuilbeufHuawei/draft-collected-data-manifest>