A Data Manifest for Contextualized Telemetry Data

draft-claise-opsawg-collected-data-manifest-06

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

IETF 116, OPSAWG

Goal & Problem Statement

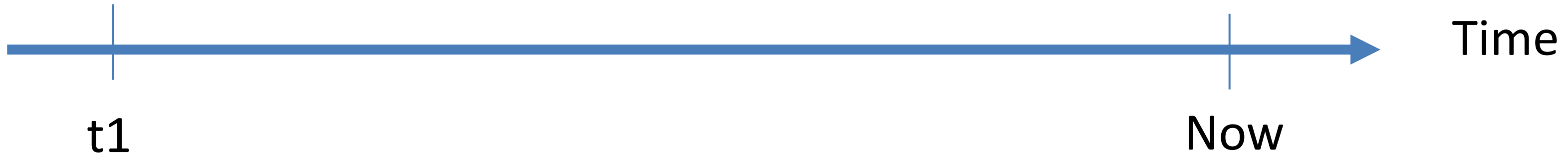
- Goal is not to expose new information via YANG but rather to define what needs to be kept as metadata (or Data Manifest) to ensure that the data can still be interpreted correctly even:
  - if the source device is not accessible (from the collection system)
  - If the source device has been updated 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

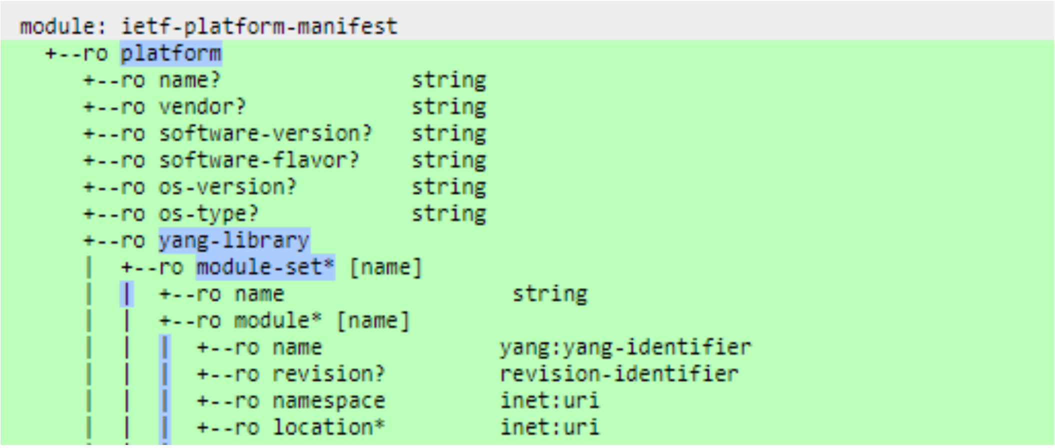
- Data Manifest composed of 2 YANG models for storing the context:
  - Platform Manifest: part of the Data Manifest that completely characterizes the platform producing the data.
  - Data Collection Manifest: part of the Data Manifest that completely characterizes how and when the telemetry was metered.
- "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: Introduction

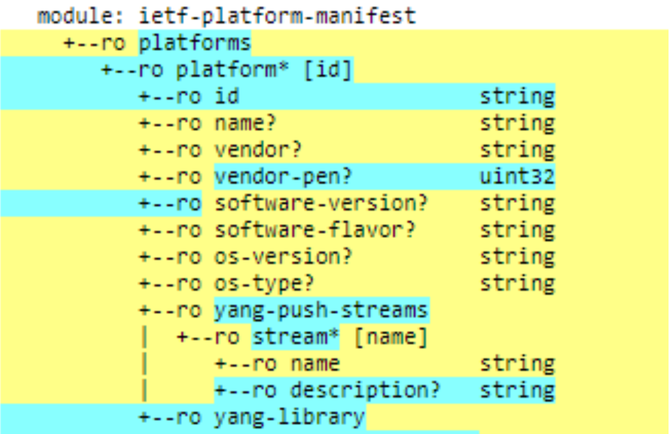
- Rewrote and clarified
- Introduces 3 identified use cases:
  - Network Analytics
  - New device onboarding
  - Data mesh principles in networking

- New “Operation Considerations” section for clarifying role of the current draft:
  - Storing data outside of the device
  - Transition period: manifest populated by collector while device support is built

Changes: Platform Manifest



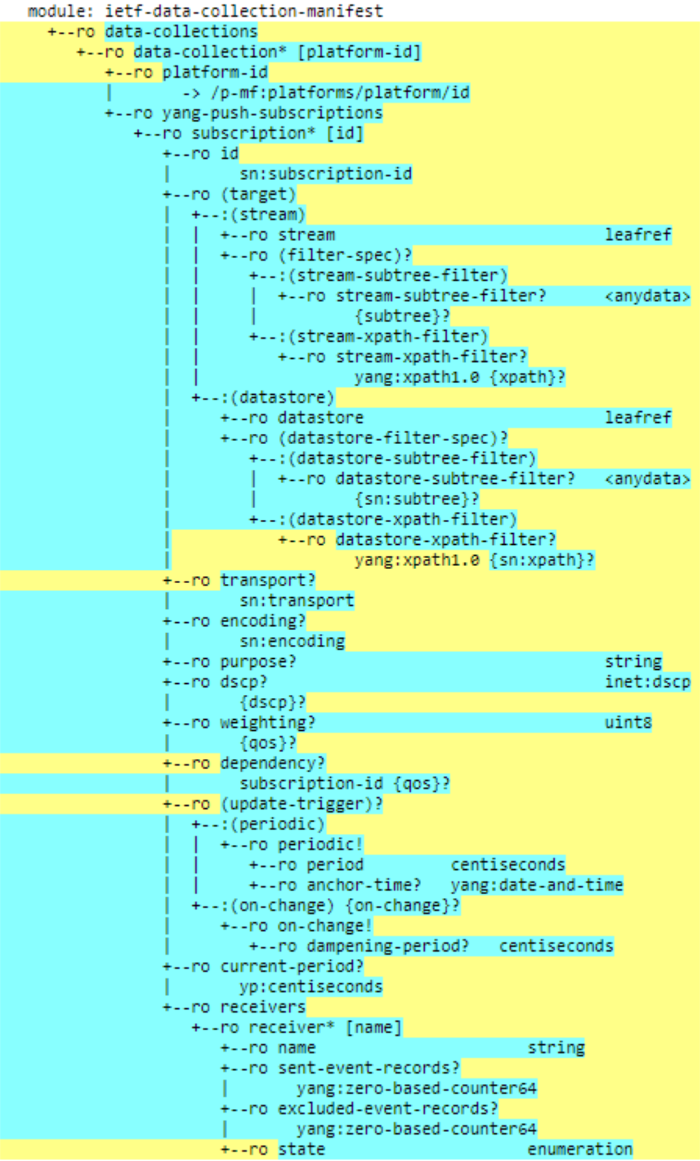
```
module: ietf-platform-manifest
  +--ro platform
     +--ro name?                string
     +--ro vendor?              string
     +--ro software-version?    string
     +--ro software-flavor?     string
     +--ro os-version?          string
     +--ro os-type?             string
     +--ro yang-library
     |  +--ro module-set* [name]
     |  |  +--ro name                   string
     |  |  +--ro module* [name]
     |  |  |  +--ro name               yang:yang-identifier
     |  |  |  +--ro revision?          revision-identifier
     |  |  |  +--ro namespace          inet:uri
     |  |  |  +--ro location*          inet:uri
```



```
module: ietf-platform-manifest
  +--ro platforms
     +--ro platform* [id]
        +--ro id                   string
        +--ro name?                string
        +--ro vendor?              string
        +--ro vendor-pen?          uint32
        +--ro software-version?    string
        +--ro software-flavor?     string
        +--ro os-version?          string
        +--ro os-type?             string
        +--ro yang-push-streams
        |  +--ro stream* [name]
        |     +--ro name           string
        |     +--ro description?   string
        +--ro yang-library
```

- Network view: top-level container is now a list of ‘platform’ indexed by an `id`.
  - Clarify that we represent data for the network, even if we reuse device modules
  - Platform id left open, suggestion: the ‘sysname’ from draft-tgraf-netconf-notif-sequencing
- Alternative way of specifying vendor: `vendor-pen` from IANA PEN
- Removed dependencies on yang-packages causing YANG errors
- Added list of streams (from RFC8639) supported by the platform for YANG push

Changes: Data Collection Manifest



```
module: ietf-data-collection-manifest
  +--ro data-collections
     +--ro data-collection* [platform-id]
        +--ro platform-id
        |       -> /p-mf:platforms/platform/id
        +--ro yang-push-subscriptions
           +--ro subscription* [id]
              +--ro id
              |       sn:subscription-id
              +--ro (target)
              |  +--:(stream)
              |  |  +--ro stream                            leafref
              |  |  +--ro (filter-spec)?
              |  |     +--:(stream-subtree-filter)
              |  |     |  +--ro stream-subtree-filter?      <anydata>
              |  |     |          {subtree}?
              |  |     +--:(stream-xpath-filter)
              |  |        +--ro stream-xpath-filter?
              |  |                yang:xpath1.0 {xpath}?
              |  +--:(datastore)
              |     +--ro datastore                         leafref
              |     +--ro (datastore-filter-spec)?
              |        +--:(datastore-subtree-filter)
              |        |  +--ro datastore-subtree-filter?   <anydata>
              |        |          {sn:subtree}?
              |        +--:(datastore-xpath-filter)
              |           +--ro datastore-xpath-filter?
              |                   yang:xpath1.0 {sn:xpath}?
              +--ro transport?
              |       sn:transport
              +--ro encoding?
              |       sn:encoding
              +--ro purpose?                                string
              +--ro dscp?                                   inet:dscp
              |       {dscp}?
              +--ro weighting?                              uint8
              |       {qos}?
              +--ro dependency?
              |       subscription-id {qos}?
              +--ro (update-trigger)?
              |  +--:(periodic)
              |  |  +--ro periodic!
              |  |     +--ro period         centiseconds
              |  |     +--ro anchor-time?   yang:date-and-time
              |  +--:(on-change) {on-change}?
              |     +--ro on-change!
              |        +--ro dampening-period?   centiseconds
              +--ro current-period?
              |       yp:centiseconds
              +--ro receivers
                 +--ro receiver* [name]
                    +--ro name                      string
                    +--ro sent-event-records?
                    |       yang:zero-based-counter64
                    +--ro excluded-event-records?
                    |       yang:zero-based-counter64
                    +--ro state                     enumeration
```

- Network view: as for platform id
  - Leaf refs to platform:
    - platform-id
    - stream
    - datastore
- Switch to YANG-push instead of MDT
  - Reuse modelling from RFC8639 and RFC8641
  - Index by subscription-id is sufficient (index by path not needed anymore)
- Additional field: ‘current-period’ between two updates, to indicate when the collection period is larger than requested (device overload)

Open Questions

- Identified Improvements:
  - Data integrity
- Open Questions:
  - Reuse device model: YANG issues: so far importing groupings and copy-pasting when no groupings are there or have a leafref that cannot be resolved in the new model. How to do this properly? Deviations? Same issue with yang-library import.
  - Handle mis-collections ? More Counters ?

Conclusion

- New draft version (Thanks Med & Tianran & Ignacio & Joe for your review)
- Do you recognize the problem statement?
- We would like to request for WG adoption

Feedback, suggestions, issues, PRs:
https://github.com/JeanQuilbeufHuawei/draft-collected-data-manifest