Weaving YANG into a Network Data Fabric

IETF 116, NMRG
YANG and the Data Fabric Paradigm

- Advanced monitoring requires combining YANG data
  - Scattered YANG data sources, both at device and service levels of abstraction
  - Multi-vendor networks, with different YANG data models and telemetry protocols
  - Need for data infrastructure to collect, integrate, and expose all these data from the network
  - Other data sources (think cloud-edge and NFV and…)
- Data Fabric architecture provides a unified view of integrated data
  - Graphs to abstract consumers from the underlying complexities of the data source
  - Standard, secure interfaces for interacting with data

The goal: Implement a data fabric solution integrating YANG monitoring data
Any Standards for Data Fabric? Enter NGSI-LD

ETSI CIM ISG defines the NGSI-LD protocol, which is composed of two parts:

**NGSI-LD Information Model**
- Based on the labelled property graph (LPG) model
- Semantic annotations based on Semantic Web standards (RDF, OWL)
- Serialized using JSON-LD

**NGSI-LD API**
- REST-based API
- Context information management
- Queries & subscriptions
- Temporal evolution
- Distributed & federated architectures
Mapping YANG Data Nodes on the NGSI-LD Metamodel

- YANG container -> NGSI-LD Entity
- YANG leaf -> NGSI-LD Relationship
- YANG leaf-list -> NGSI-LD Property
- YANG list
Starting with a Simple Use Case: Interfaces

```
module: ietf-interfaces
  +++ rw interfaces
    +++ rw interface [name]
      +-- rw name            string
      +-- rw description?    string
      +-- rw type            identityref
      +-- rw enabled?
      +-- rw link-up-down-trap-enable?
        +--rw admin-status   enumeration {if-mib}?
        +--rw oper-status    enumeration
        +--rw last-change?   yang:date-and-time
        +--rw if-index       int32 {if-mib}?
        +--rw phys-address? .yang:phys-address
        +--rw higher-layer-if* interface-ref
        +--rw lower-layer-if* interface-ref
        +--rw speed?         yang:auge64
      +--rw statistics
        +--rw discontinuity-time yang:date-and-time
        +--rw in-octets?      yang:counter64
        +--rw in-unicast-pkts? yang:counter64
        +--rw in-broadcast-pkts? yang:counter64
        +--rw in-multicast-pkts? yang:counter64
        +--rw in-discards?    yang:counter32
        +--rw in-errors?      yang:counter32
        +--rw in-unknown-protos? yang:counter32
        +--rw out-octets?     yang:counter64
        +--rw out-unicast-pkts? yang:counter64
        +--rw out-broadcast-pkts? yang:counter64
        +--rw out-multicast-pkts? yang:counter64
        +--rw out-discards?   yang:counter32
        +--rw out-errors?     yang:counter32
```

Tree representation of ietf-interfaces@2018-02-20.yang
Managing YANG Configuration vs Operational Data

- NGSI-LD defines the `datasetId` feature:
  - Entity attributes can have multiple values provided by different data sources
  - Represented as a sub-property that uniquely identifies the source
- Feature allows for creating graph projections using the same data model
  - Avoid duplication of data
- **Proposal →** Leverage `datasetId` feature to identify YANG datastores (pre/post-NMDA)
Data Fabric Prototype

- **Containerlab** for network virtualization
  - Based on Docker containers
- Testing devices that implement YANG Push and IETF modules (NMDA ideally)
  - Cisco CSR1000v
  - Looking to include more vendors
- Two custom components
  - YANG Push collector, leveraging the `ncclient` library
  - YANG to NGSI-LD translator
A Few Research Challenges to Start With

• Address the whole YANG scope
  • Refine mapping rules with more complex YANG data models
  • Connect YANG data from device and service levels → The network map approach

• Mapping rules focused on NGSI-LD information model
  • Could be the basis for mappings to LPG (e.g., Neo4j) and RDF (generation of OWL ontologies from YANG data models)

• Align prototype development with YANG native onboarding
  • [GitHub Link]
  • Data fabric as an enabler of data mesh approach

• Collaboration with ETSI CIM ISG
  • Details of mapping rules in new group report

• Going beyond monitoring data
  • Integration of configuration data → The other half of the closed-loop

*This Communication is part of a project that has received funding from the European Union’s Horizon 2020 research and innovation program under grant agreement No 883335.*