Designing a network measurement ontology for a semantically driven architecture

Alfredo Salvador, Álvaro Katsu
Jorge E. López de Vergara, Javier Aracíl
HPCN UAM SPAIN

Agenda

- Objectives
- Moment architecture
- Moment Ontologies
- Moment Ontologies v2
  - Units
  - Data
  - Metadata
- Mapping process
- Conclusions and Future Work

Objectives

- Difficult adoption of the ontology in the MOMENT architecture
- Modules not included in the plan now need to benefit from semantic technologies
- Anonymization ontology must remain unchanged
- Complicated mapping process
- Some structural complains about the ontology hierarchy

Moment architecture

- MOMENT is based on SOA, each user query goes through a chain of Services working on data from several sources.
- MOMENT intends to merge very diferent measurement datasources each one with its own vocabulary
- MOMENT Ontologies try to fix this heterogeneity to provide the user a transparent and independent from sources platform

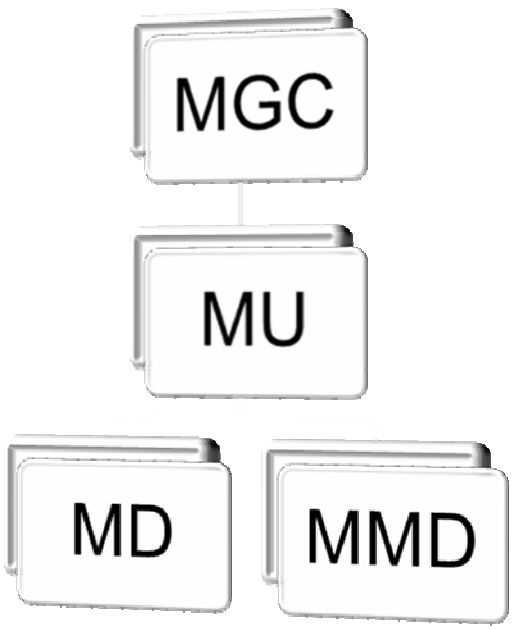
Moment Ontologies v1

- Hierarchy of network measurements based on the Active / Pasive division.
- Data ontology based on the previous work from OGF NMWG
- Metadata ontology based on DatCat's object types
- Strict classification of measurements under the hierarchy -> Mapping process dificult ☹

Moment Ontologies v1

- MOMENT Partner's complains and necessities:
  - Better correspondence between the metadata ontology and CAIDA's DatCat object types.
  - Definition of the necessary units for measurements: bit, Byte, IP ?, MAC ?, etc.
  - Redefine the Data ontology to a metric oriented ontology and provide a flexible way of mapping.
  - Define in the DataMetadata object information about the result columns names and types from each data source so GUI and GMF don't need to Query.

Moment Ontologies v2



The ontology is divided in four smaller ontologies:
- Moment General Concepts (***MGC***)
- Moment Units (***MU***)
- Moment Data (***MD***)
- Moment Metadata (***MMD***)

MGC contains the common information about network measurements such as Protocols, time stamps and Locations

Units (wich will be part of MGC) defines the network measurement units needed for the result columns

Moment Data defines the stored measurements in the datasource (usually Relational tables) and Moment Metadata describes those datasources and their datasheets.

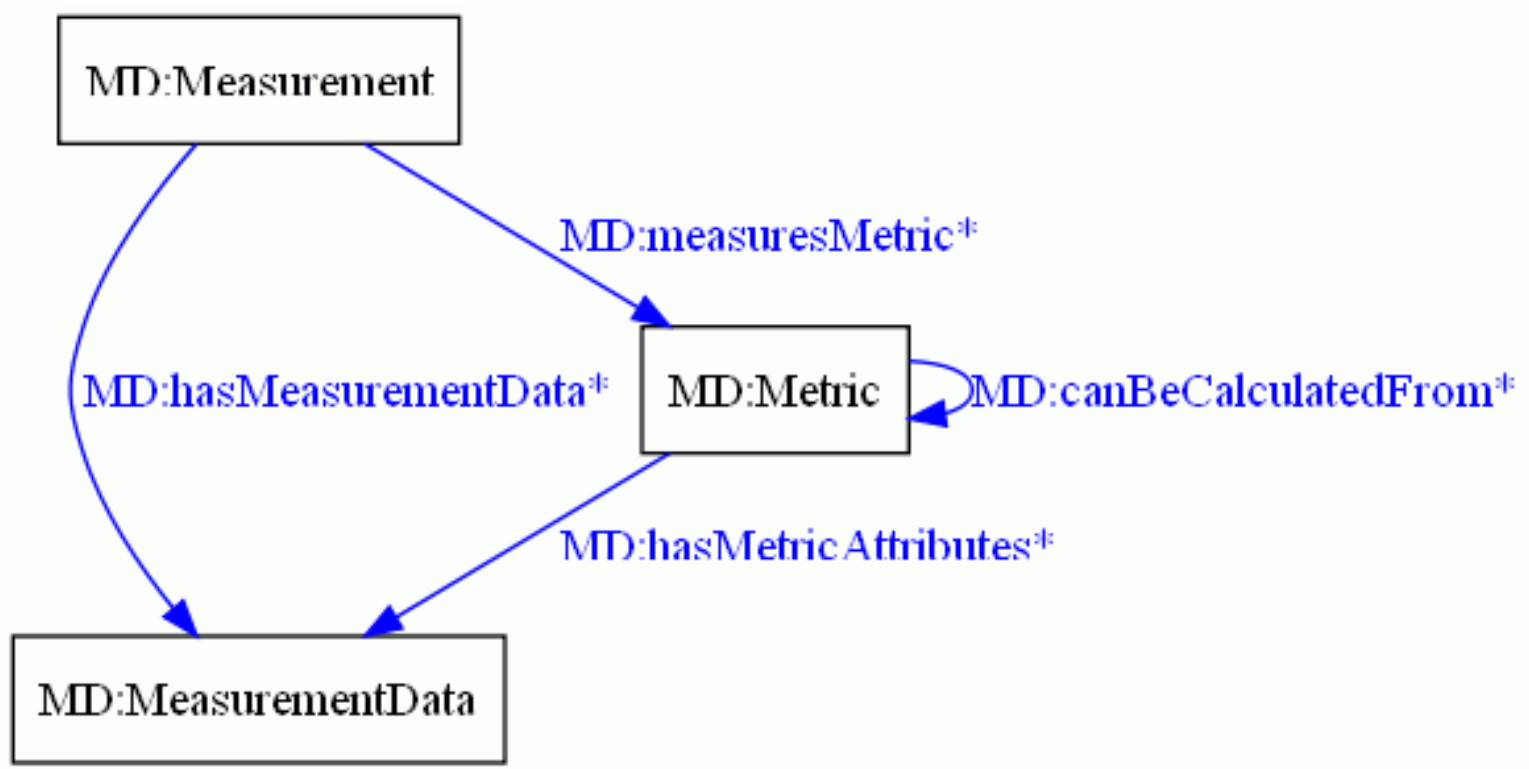
Units Ontology

- We adopted NASA units ontology (http://sweet.jpl.nasa.gov/ontology/units.owl)
  - Adding support for Computer Sciences units like bit and Byte
  - Metric and Binary prefixes: Mega & Mebi
  - IP and MAC as units for column results
  - Transformation functions for numeric and non-numeric values.

Units Ontology

- SDTS (Semantic Data Transformation Service) benefits from the new version of the ontology in multiple ways:
  - With the new version of the ontology now SDTS is able to retrieve for each column in the results the unit and provide automatic unit conversion.
  - Also datasources can use different units and SDTS will still be able to transform to a common unit

Data Ontology

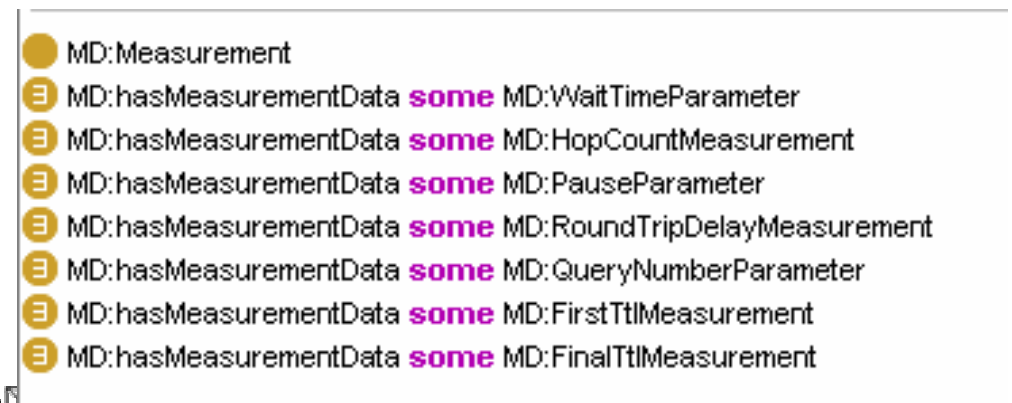


Data Ontology

- More flexible mappings are possible:
  - Measurements represent tables
  - MeasurementData represent columns
  - Metrics join the high level concepts from network measurements to particular measurement realizations
- Passive/Active hierarchy is replaced with a relaxed classification based on *“what information the measurement has”*

Common measurements

- E.g. Traceroute: Well known measurement techniques such as Traceroute are described as subclasses of MD:Measurement
  - Not adding more properties but adding constraints to the MeasurementData the measurement should have.

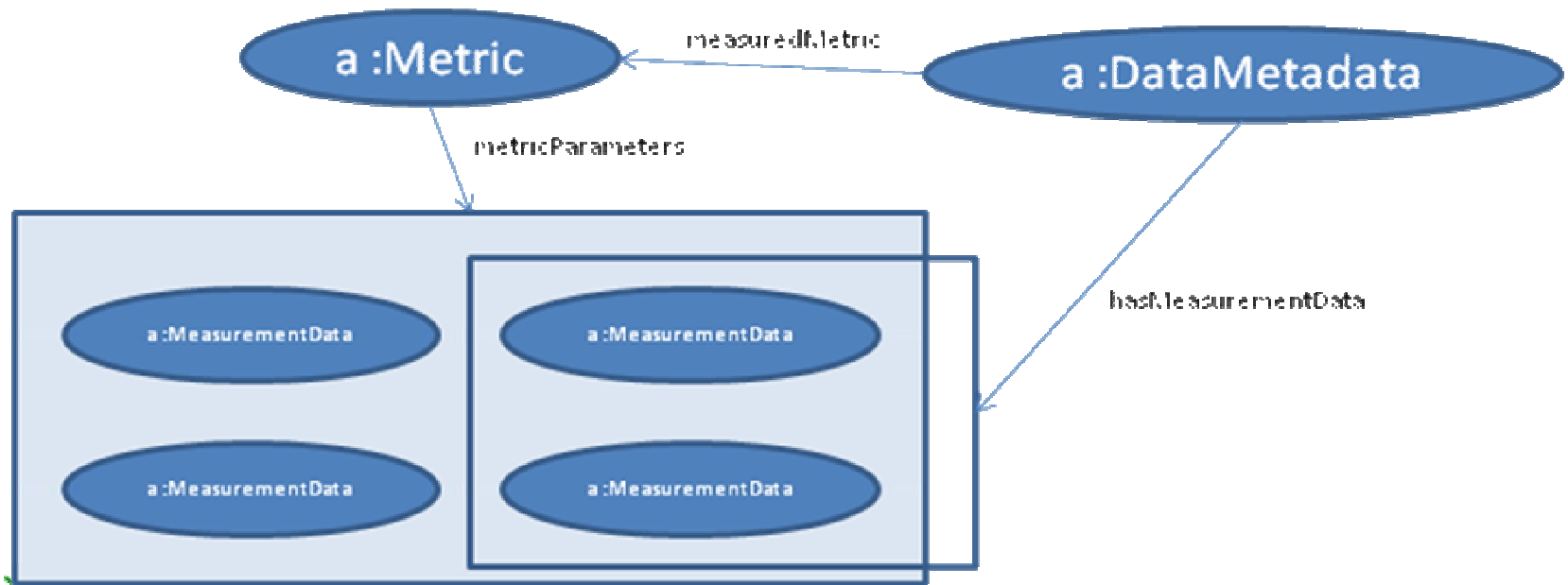


Metadata Ontology

- The Metadata ontology has been better aligned with DatCat's object types
  - This allows SMR to import semi-automatically all the information from DatCat
  - *Moment will have at least the same information as DatCat*
- Other concepts from DatCat such as publication are not used in MOMENT

Metadata Ontology

- Also the ontology provides ways to define the structure of each datasource results.



Mapping process

- Now the mapping process for a datasource is easier, and even more for relational databases if mapped using D2R.
  - Tables are mapped to MD:Measurement
  - Columns are mapped to some subclass of MD:MeasurementData wich represents the value measured.
    - Also the unit of the measurement can be specified via a fixed property in the configuration file.
  - Metrics are asigned to measurements either manually by the datasource mantainer or automatically through a inference engine.

Mapping process

- With D2R, mappings are done through configuration files:

```
# Table NtpGps_measure
map3:NtpGps_measure a d2rq:ClassMap;
      d2rq:dataStorage map3:database;
      d2rq:uriPattern "NtpGps_measure/@@NtpGps_measure.Id@@";
      d2rq:class vocab:NtpGps_measure;
      .
```

- Some parts of those files need to match the ontology for the system to understand it
- The process can be extrapolated to non-relational data sources using CVS configuration files

Conclusions and Future Work

- The new version of the ontology provides every MOMENT partner the tool they needed to use semantic technologies in their module.
- The mapping process now it’s very simple, making new users willingly to map their datasources. Also the standarization of the ontology will help in this point.
- Datasources are described in terms of the ontology wich is usefull for Metadata retrieval and for query federation. Future work in this field is planned because well known problems of query federation need to be solved for the final MOMENT architecture.

THANKS!