Data Type Registries WG Progress Reports
RDA-P9

The ePIC Data Type Registry

Ulrich Schwardmann

Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen (GWDG)
Am Fassberg, 37077 Göttingen
ulrich.schwardmann [at] gwdg.de

06 April 2017, Barcelona
Status of the ePIC DTR

- [http://dtr.pidconsortium.eu/](http://dtr.pidconsortium.eu/)

- number of types
  - BasicInfoTypes: 150
  - InfoTypes: 89
  - Types: 268
    - including schema, user, ...

- quality of service assurance:
  - monitoring, backup
  - puppetized
hierarchical type definitions

- concept of *basic PID info types* and *PID info types*
  - makes it easy to define new types
  - to provide types for an adaption of the collections-WG recommendations: 1h
  - most complex data type currently: DataCite Mandatory MD properties (depth 5)

- in consequence:
  - possibility to derive schemas for the type values
  - automated server side schema derivation
in preparation, candidate, approved and deprecated

- first candidates defined
- migration from *in preparation* to *candidate* is non-trivial
  - because two different DTRs are involved
  - and because of the hierarchical concept
  - all subtypes need to be candidates
  - the subtype relation need to be automatically transformed
International System Of Units
SI by NIST

- all basic and derived units and prefixes defined
- because they are always numbers, they are all basic PID info types with format number
  - challenge: a way to represent units by other units
    - this needs an entry, where the derivation is expressed
    - and a way to represent constants as types
    - problem here is that types are variables by nature
Current Implementation of SI at ePIC DTR

Energy derived unit

```
{
  "name": "energy SI",
  "provenance": {
    "creationDate": "2017-04-04T17:14:47.706Z",
    "lastModificationDate": "2017-04-04T17:15:03.557Z",
    "contributors": [
      {
        "identifiedUsing": "Text",
        "name": "Ulrich Schwardmann",
        "details": "GWDG"
      }
    ],
    "validationSchema": "{\"definitions\": {}, \"$schema\": \"http://json-schema.org/draft-04/schema\"",
    "standards": [
      {
        "details": "http://physics.nist.gov/cuu/Units/units.html",
        "natureOfApplicability": "depends",
        "name": "International System of Units",
        "issuer": "NIST"
      },
      {
        "natureOfApplicability": "depends",
        "name": "21.T11148/7e2c0ec2e74e3ce2f85d",
        "issuer": "DTR"
      },
      {
        "natureOfApplicability": "depends",
        "name": "21.T11148/0a1af7f849bc91469df",
        "issuer": "DTR"
      }
    ],
    "identifier": "21.T11148/8f6a2d356fbb80a1478f",
    "properties": [
      {
        "dataType": "number",
        "symbol": "J",
        "expression": "Measurement Unit",
        "definedBy": "21.T11148/7e2c0ec2e74e3ce2f85d * 21.T11148/0a1af7f849bc91469df",
        "value": "joule"
      }
    ],
    "description": "SI derived unit for energy, work or quantity of heat by joule"
  }
}
```
use a computer algebra approach for unit transformation

- PIDs are variables for CA
- use CA on these variables, and compute
  - the fields of "value" then yield to equation like
    \[ \text{energy} = \text{force} \times \text{length} \]
  - the fields symbol then yield to dimension of the unit like
    \[ J = N \times m \]
    \[ J = m^{**2} \times kg \times s^{**-2} \]
- a unit converter based on these types is a very feasible task