

First International Workshop on Reproducible Open Science
Hannover, Germany, September 9, 2016

From Data to Machine Readable Information Aggregated in Research Objects

Markus Stocker

PANGAEA / MARUM
University of Bremen
Germany

<http://orcid.org/0000-0001-5492-3212>
@envinf



marum
Center for Marine
Environmental Sciences



 **Universität Bremen**



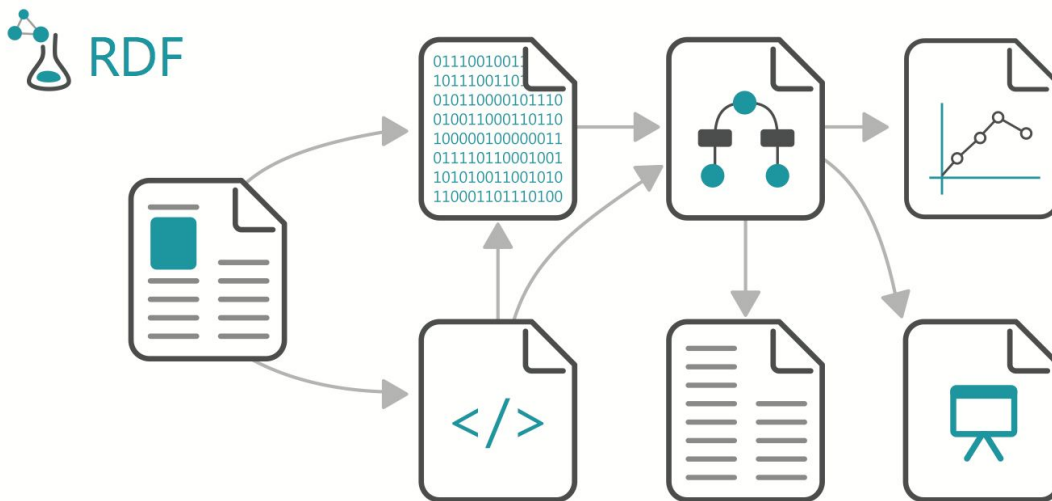
PLUS
ENVRI

Introduction

- Data interpretation is key in scientific investigations
- Process with data as input and information as output
- Data are uninterpreted symbols, e.g. sensor observation values
- Information are interpreted data, for their meaning in a real-world context
- Record information resulting in data interpretation

Research Object

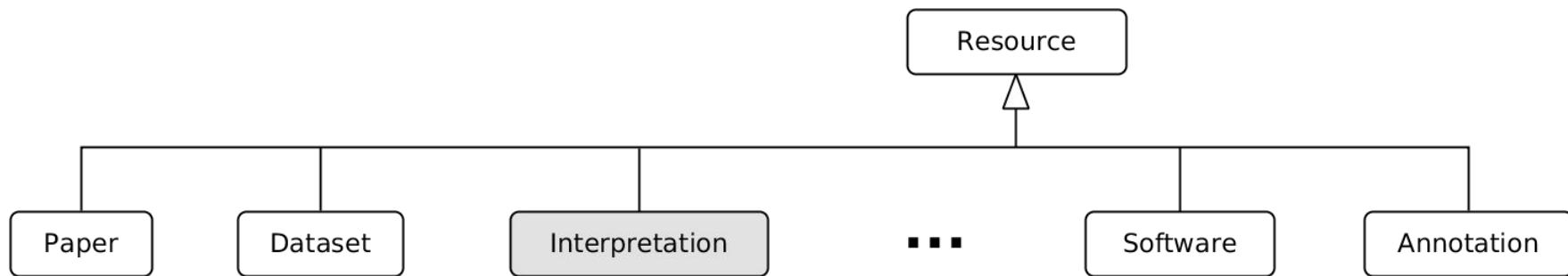
- “Semantically rich aggregations of resources that bring together data, methods and people in scientific investigations” (Bechhofer et al., 2013)



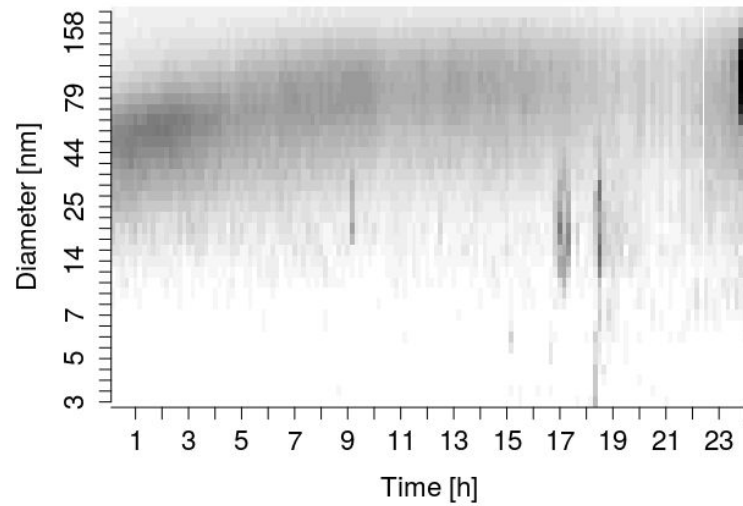
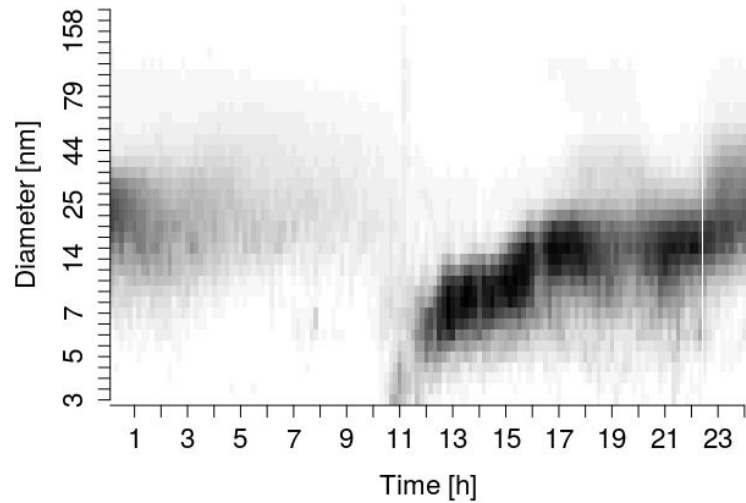
Proposal

- Extend the Research Object model
- Additional Resource type called *Interpretation*
- Instances represent information resulting from data interpretation
- Instances are machine readable

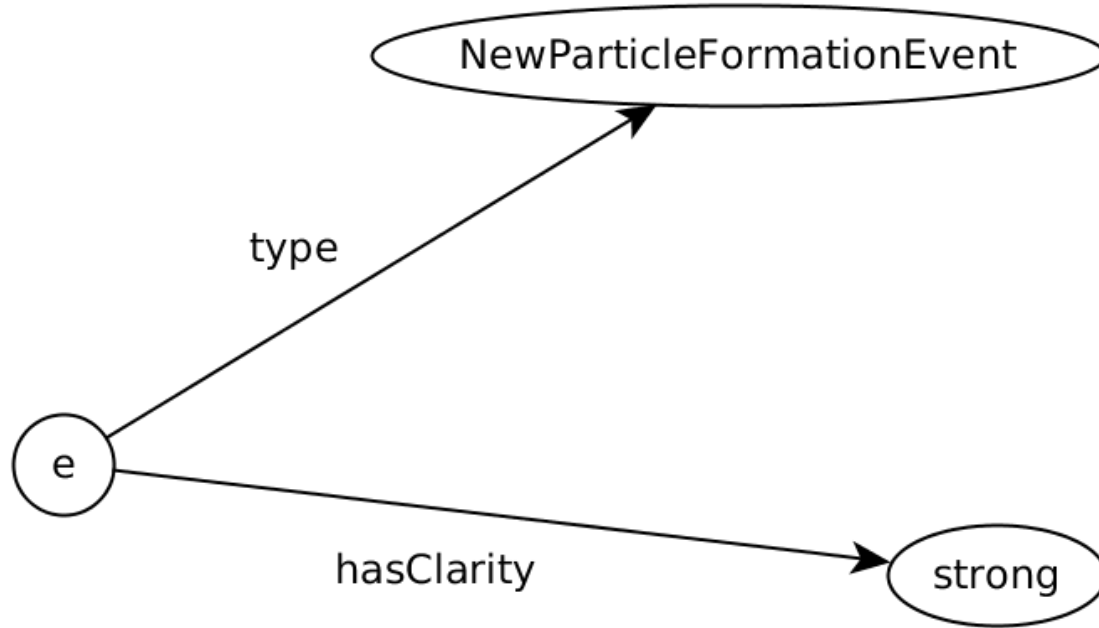
Proposed Extension



Application



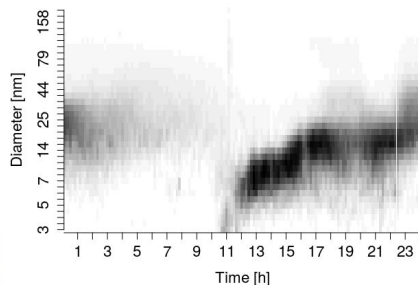
Application



Application

```
@prefix qb: <http://purl.org/linked-data/cube#>.
@prefix ex: <http://example.org#>.
@prefix dimension: <http://purl.org/linked-data/sdmx/2009/dimension#>.

ex:c819b0f-24b5-449a-8d66-54cbb17afe7
  a qb:Observation ;
  ex:d108a2 <"4301.86"^^http://www.w3.org/2001/XMLSchema#double>;
  ex:d108b3 <"146.488"^^http://www.w3.org/2001/XMLSchema#double>;
  ex:d126a2 <"5765.31"^^http://www.w3.org/2001/XMLSchema#double>;
  ex:d126b3 <"134.662"^^http://www.w3.org/2001/XMLSchema#double>;
  ex:d126c2 <"6762.73"^^http://www.w3.org/2001/XMLSchema#double>;
  ex:d126d3 <"119.767"^^http://www.w3.org/2001/XMLSchema#double>;
  ex:d141e2 <"7563.71"^^http://www.w3.org/2001/XMLSchema#double>;
  ex:d141e3 <"101.166"^^http://www.w3.org/2001/XMLSchema#double>;
  ex:d158a2 <"8196.98"^^http://www.w3.org/2001/XMLSchema#double>;
  ex:d158b3 <"84.262"^^http://www.w3.org/2001/XMLSchema#double>;
  ex:d178e2 <"7767.8"^^http://www.w3.org/2001/XMLSchema#double>;
  ex:d178e3 <"74.092"^^http://www.w3.org/2001/XMLSchema#double>;
  ex:d208e2 <"6753.22"^^http://www.w3.org/2001/XMLSchema#double>;
  ex:d208b3 <"57.6419"^^http://www.w3.org/2001/XMLSchema#double>;
  ex:d208c2 <"6467.31"^^http://www.w3.org/2001/XMLSchema#double>;
  ex:d251e2 <"3086.22"^^http://www.w3.org/2001/XMLSchema#double>;
  ex:d251e3 <"2156.81"^^http://www.w3.org/2001/XMLSchema#double>;
  ex:d316e2 <"251.075"^^http://www.w3.org/2001/XMLSchema#double>;
  ex:d316e3 <"1789.69"^^http://www.w3.org/2001/XMLSchema#double>;
  ex:d355e2 <"158.213"^^http://www.w3.org/2001/XMLSchema#double>;
  ex:d355e3 <"1184.19"^^http://www.w3.org/2001/XMLSchema#double>;
  ex:d398e2 <"265.944"^^http://www.w3.org/2001/XMLSchema#double>;
  ex:d398e3 <"625.349"^^http://www.w3.org/2001/XMLSchema#double>;
  ex:d447e1 <"415.541"^^http://www.w3.org/2001/XMLSchema#double>;
  ex:d447e2 <"374.991"^^http://www.w3.org/2001/XMLSchema#double>;
  ex:d501e1 <"508.331"^^http://www.w3.org/2001/XMLSchema#double>;
  ex:d501e2 <"723.747"^^http://www.w3.org/2001/XMLSchema#double>;
  ex:d526e1 <"1006.54"^^http://www.w3.org/2001/XMLSchema#double>;
  ex:d526e2 <"225.007"^^http://www.w3.org/2001/XMLSchema#double>;
  ex:d616e1 <"1663.32"^^http://www.w3.org/2001/XMLSchema#double>;
  ex:d616e2 <"173.586"^^http://www.w3.org/2001/XMLSchema#double>;
  ex:d708e1 <"2761.95"^^http://www.w3.org/2001/XMLSchema#double>;
  ex:d708e2 <"175.845"^^http://www.w3.org/2001/XMLSchema#double>;
  ex:d794e1 <"3411.52"^^http://www.w3.org/2001/XMLSchema#double>;
  ex:d794e2 <"158.71"^^http://www.w3.org/2001/XMLSchema#double>;
  ex:d919e1 <"3207.95"^^http://www.w3.org/2001/XMLSchema#double>;
  ex:d919e2 <"148.043"^^http://www.w3.org/2001/XMLSchema#double>;
  qb:dataSet
  dimension:timePeriod <"2011-03-26T08:29:00"^^http://www.w3.org/2001/XMLSchema#dateTime> .
```



```
@prefix ex: <http://example.org#>.
@prefix dimension: <http://purl.org/linked-data/sdmx/2009/dimension#>.
@prefix qb: <http://purl.org/linked-data/cube#>.

ex:5df2e06e-5288-4cf9-a04e-6e8c204646b1
  a qb:Observation ;
  ex:p01 <"182721.205992693"^^http://www.w3.org/2001/XMLSchema#double>;
  ex:p02 <"91342.053033951"^^http://www.w3.org/2001/XMLSchema#double>;
  ex:p03 <"42964.8233814576"^^http://www.w3.org/2001/XMLSchema#double>;
  ex:p04 <"26473.321306687"^^http://www.w3.org/2001/XMLSchema#double>;
  ex:p05 <"13416.02504344"^^http://www.w3.org/2001/XMLSchema#double>;
  ex:p06 <"9393.90646981048"^^http://www.w3.org/2001/XMLSchema#double>;
  ex:p07 <"8465.7579595562"^^http://www.w3.org/2001/XMLSchema#double>;
  ex:p08 <"6732.99213687614"^^http://www.w3.org/2001/XMLSchema#double>;
  ex:p09 <"5888.09222677874"^^http://www.w3.org/2001/XMLSchema#double>;
  ex:p10 <"4851.9029806067"^^http://www.w3.org/2001/XMLSchema#double>;
  ex:p11 <"4689.78576752719"^^http://www.w3.org/2001/XMLSchema#double>;
  ex:p12 <"3529.9472878159"^^http://www.w3.org/2001/XMLSchema#double>;
  ex:p13 <"2800.15913938687"^^http://www.w3.org/2001/XMLSchema#double>;
  ex:p14 <"2336.02027821001"^^http://www.w3.org/2001/XMLSchema#double>;
  ex:p15 <"1693.5233253359"^^http://www.w3.org/2001/XMLSchema#double>;
  ex:p16 <"1426.11874696801"^^http://www.w3.org/2001/XMLSchema#double>;
  ex:p17 <"1005.00962345565"^^http://www.w3.org/2001/XMLSchema#double>;
  ex:p18 <"727.531591381707"^^http://www.w3.org/2001/XMLSchema#double>;
  ex:p19 <"597.05170688255"^^http://www.w3.org/2001/XMLSchema#double>;
  ex:p20 <"525.63284620044"^^http://www.w3.org/2001/XMLSchema#double>;
  ex:p21 <"290.00227863251"^^http://www.w3.org/2001/XMLSchema#double>;
  ex:p22 <"254.090693216949"^^http://www.w3.org/2001/XMLSchema#double>;
  ex:p23 <"231.20448747975"^^http://www.w3.org/2001/XMLSchema#double>;
  ex:p24 <"218.800935984002"^^http://www.w3.org/2001/XMLSchema#double>;
  ex:p25 <"156.975828147214"^^http://www.w3.org/2001/XMLSchema#double>;
  ex:p26 <"141.530875977794"^^http://www.w3.org/2001/XMLSchema#double>;
  ex:p27 <"129.56393016864"^^http://www.w3.org/2001/XMLSchema#double>;
  ex:p28 <"0.109348089480993"^^http://www.w3.org/2001/XMLSchema#double>;
  ex:p29 <"0.0936781322005406"^^http://www.w3.org/2001/XMLSchema#double>;
  ex:p30 <"0.057418608088215"^^http://www.w3.org/2001/XMLSchema#double>;
  ex:p31 <"0.0499170120683776"^^http://www.w3.org/2001/XMLSchema#double>;
  ex:p32 <"0.0268110452149435"^^http://www.w3.org/2001/XMLSchema#double>;
  ex:p33 <"0.0171719898374105"^^http://www.w3.org/2001/XMLSchema#double>;
  ex:p34 <"0.015455325631149"^^http://www.w3.org/2001/XMLSchema#double>;
  ex:p35 <"0.0073117555122899"^^http://www.w3.org/2001/XMLSchema#double>;
  ex:p36 <"0.00181626248315137"^^http://www.w3.org/2001/XMLSchema#double>;
  ex:p37 <"0.00208493748712496"^^http://www.w3.org/2001/XMLSchema#double>;
  qb:dataSet
  dimension:timePeriod
  dimension:timePeriod <"2011-03-26"^^http://www.w3.org/2001/XMLSchema#date> .

ex:d2 a qb:DataSet ;
  <http://www.w3.org/1999/02/22-rdf-syntax-ns#e8ee1a0>
  <https://figshare.com/s/162ee315f878a56ba8e>, <https://doi.org/10.6084/m9.figshare.3571146> .
```

```
# -----
# Author: Markus Stocker
# ORCID: 0000-0001-5492-3212
# Date: 2016-08-11
# License: CC-BY
# -----
library(Rcur1)
# https://github.com/egonw/rrdf
# Requires rJava, see:
# http://stackoverflow.com/a/2593282
library(rrdf)
library(uid)

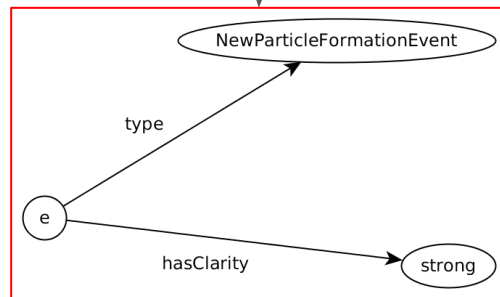
# ---- CONFIGURATION ----
Sys.setlocale("LC_TIME", "en_US.UTF-8")

host <- "http://avaa.tdata.fi"
table <- "HYV_DNPS"
from <- "2011-03-26 00:00:00.000"
to <- "2011-03-27 00:00:00.000"
quality <- "ANY"
averaging <- "NONE"
type <- "NONE"
variables <- c("d316e1","d355e1","d398e1","d447e1","d501e1","d562e1",
"d631e1","d708e1","d794e1","d891e1","d100e2","d112e2",
"d126e2","d141e2","d158e2","d178e2","d200e2","d224e2",
```

stocker16hamover-repescience-workflows (6/72 kb)

Download (672 kb)

This item is shared privately



Application

```
<>      a          ro:ResearchObject ;
         ore:aggregates ex:d1, ex:f1, ex:d2, ex:s1, ex:e ;
         dct:created   "2016-08-11"^^xsd:dateTime ;
         dct:creator   [ a foaf:Person; foaf:name "Markus Stocker" ] .

ex:d1    a          wf4ever:Dataset, qb:DataSet ;
         swrc:doi    <https://doi.org/10.6084/m9.figshare.3565635> .

ex:f1    a          wf4ever:Image ;
         swrc:doi    <https://doi.org/10.6084/m9.figshare.3567273> .

ex:d2    a          wf4ever:Dataset, qb:DataSet ;
         swrc:doi    <https://doi.org/10.6084/m9.figshare.3571146> .

ex:s1    a          wf4ever:Software ;
         swrc:doi    <https://doi.org/10.6084/m9.figshare.3571212> .

ex:e     a          ex:NewParticleFormationEvent, ro:Interpretation ;
         ex:hasClarity ex:strong .
```

Discussion

- Relevance of ontology because interpretation follows a conceptualization
- Share semantics between humans and machines
- Utilize interpretations to build models, e.g. machine learning classifiers
- Applicable also to Distributed Scholarly Compound Object (DiSCO)
- Link other PID types, e.g. ORCID iD

```
<> a ro:ResearchObject ;  
    dct:creator [ a foaf:Person; dbo:orcidId "0000-0001-5492-3212" ] .
```

Conclusion

- Data interpretations are artefacts in scientific investigations
- Record interpretations in artefact aggregations, e.g. Research Object
- Record for humans *and* machines, not just images and natural language text