

Data Modeling for QPRF/QMRF Reports

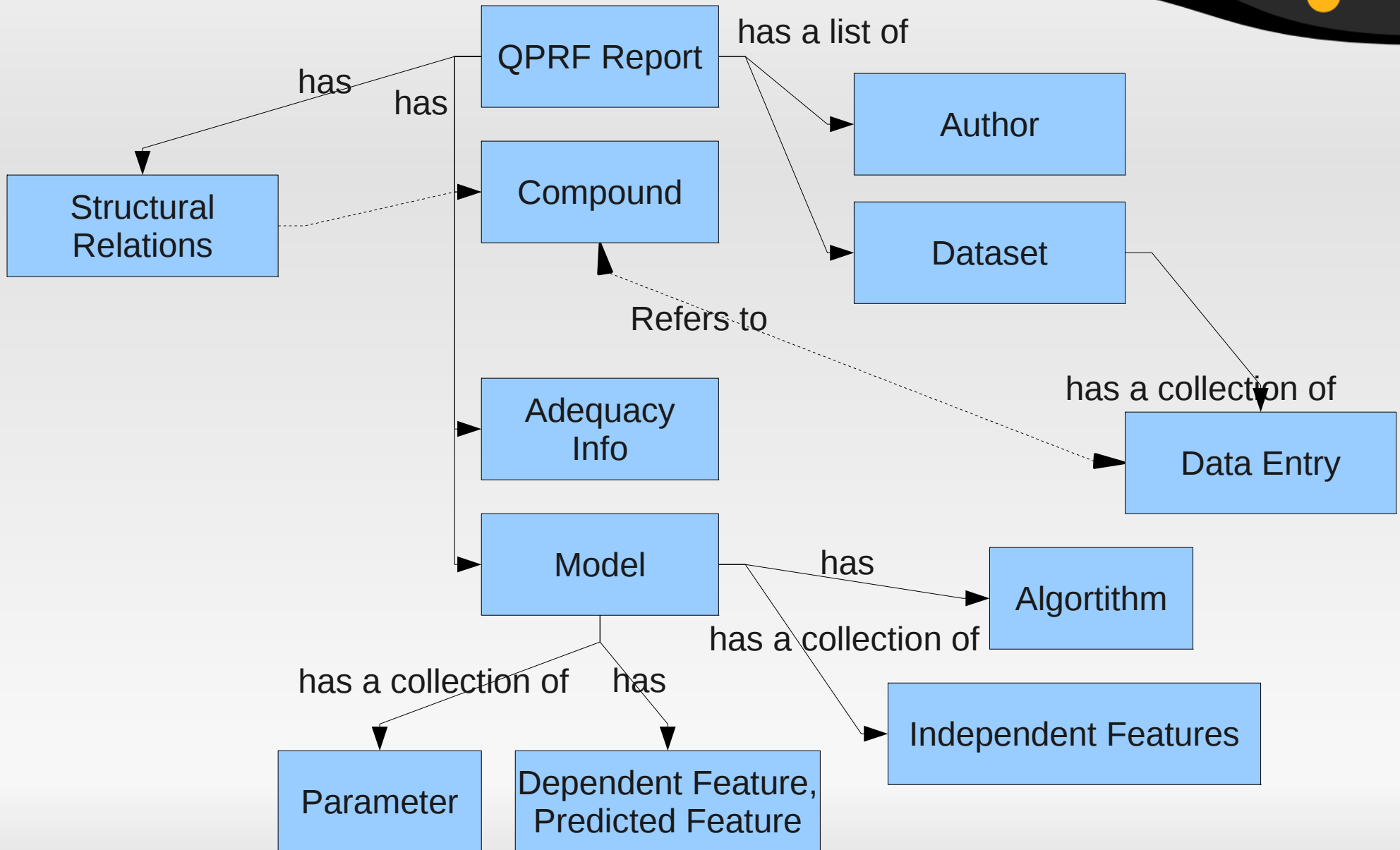


- Ontology Extensions Necessary for the modeling of QPRF (mainly) and QMRF reports.
- Illustrative Examples

Sopasakis Pantelis (NTUA)
chvng@mail.ntua.gr

Note: Where necessary, newly introduced properties and classes appear in *italics*.

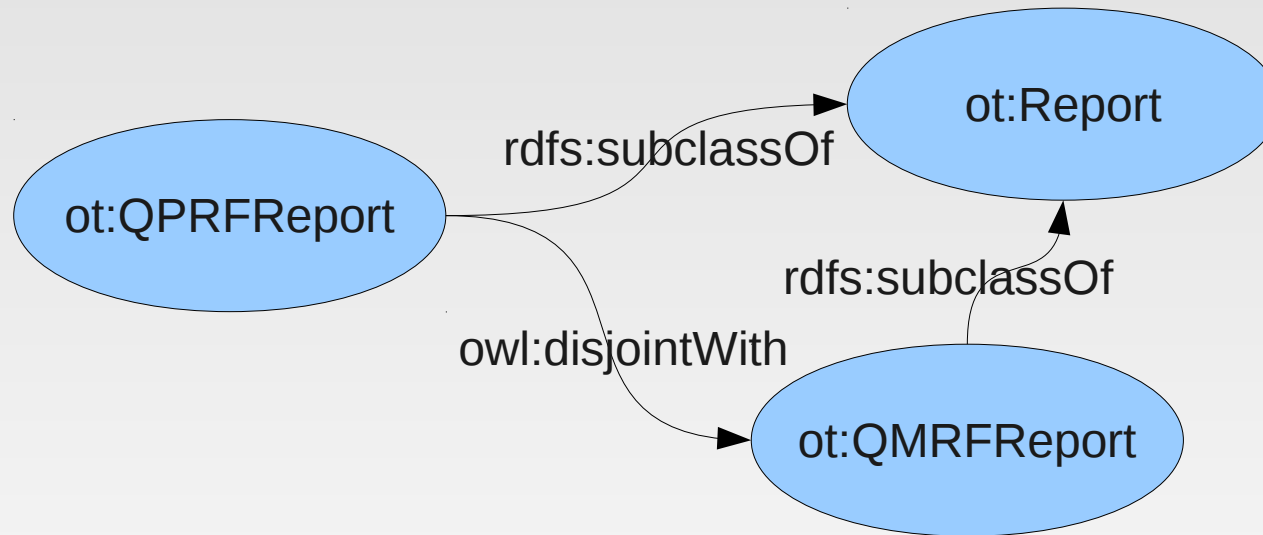
QPRF Report - Outline



QPRF and QMRF Reports



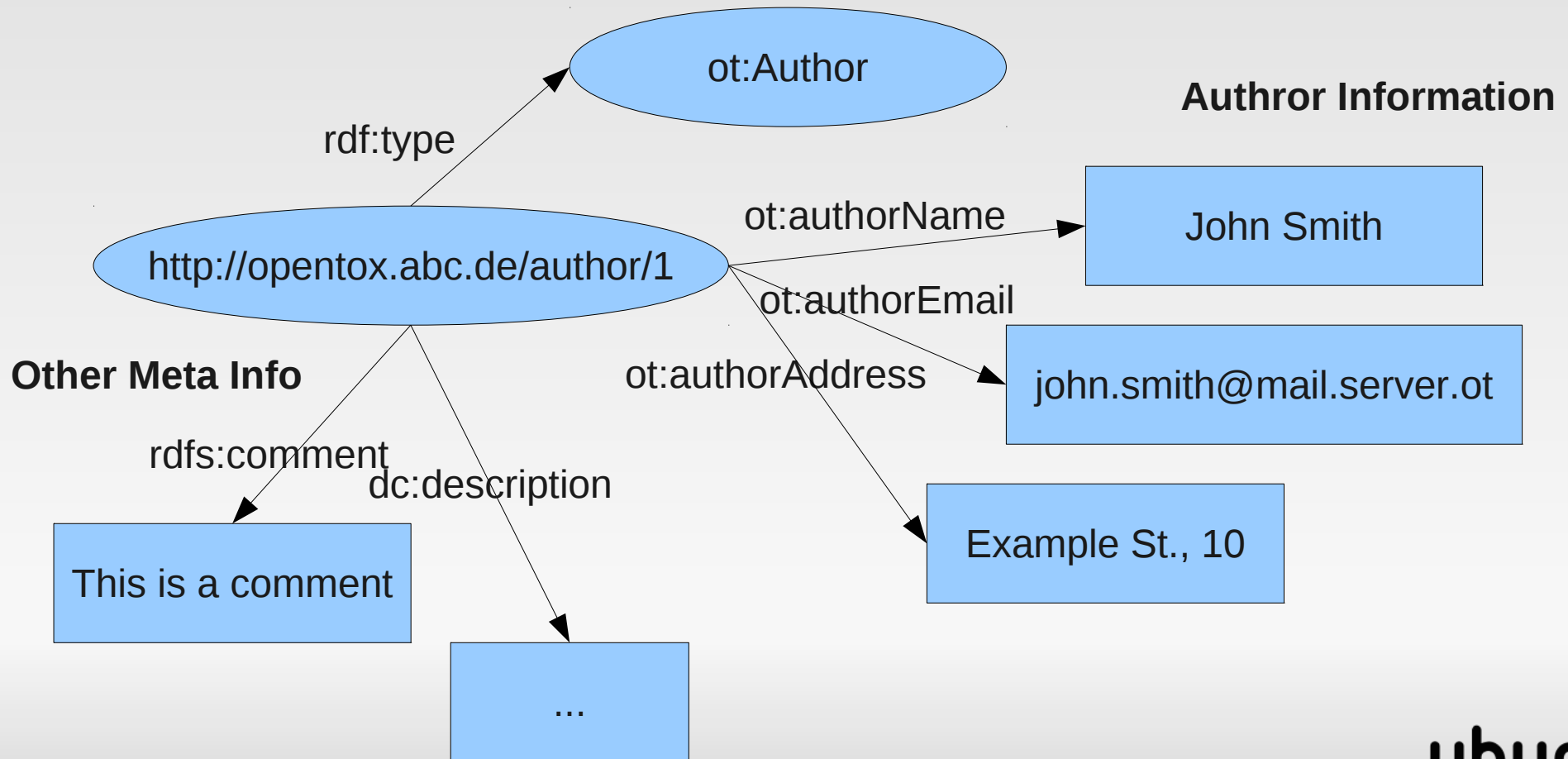
The generic ontological class *ot:Report* is subclassed by *ot:QPRFReport* and *ot:QMRFReport* for QPRF and QMRF reports respectively.



Authors...



dc:creator is just a string... An *author* should be a new ontological class (ot:Author) having a name, an email, a URL, an address, phone number, whatever else we need... Here is an example:



Ordered Lists in RDF

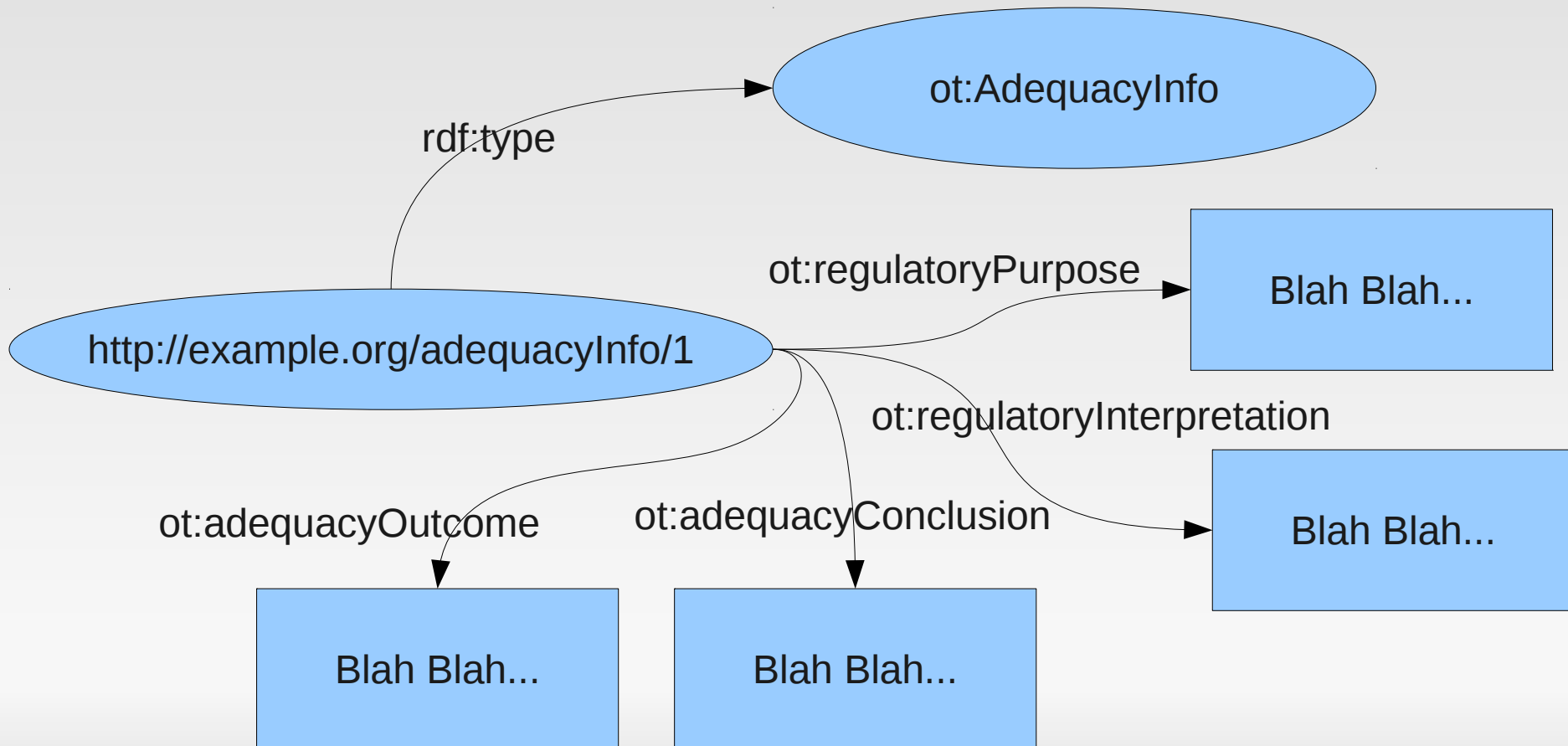


- RDF was not designed to preserve order in the information it holds. However... This is what containers are all about.
- `<rdf:Seq>` is used to describe an ordered list of values
- See section 2.15 of RDF syntax specifications
<http://www.w3.org/TR/REC-rdf-syntax/#section-Syntax-list-elements>
- List of authors should be order preserving

Adequacy Information



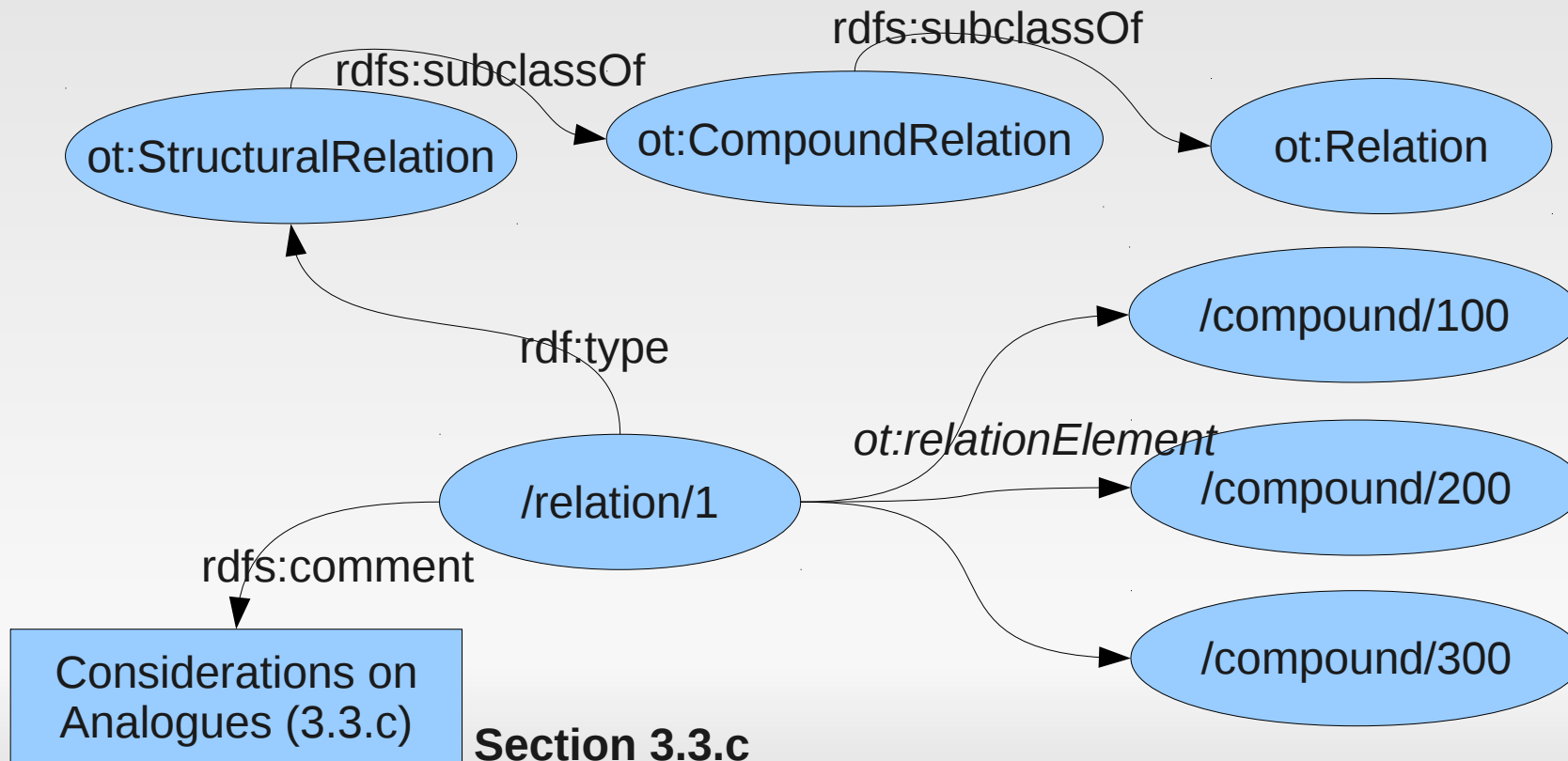
This refers to section 4 of QPRF report. Here is an example of an RDF Node about Adequacy Information found in a QPRF report...



Structural Analogues



Structural Analogues of the compound into consideration should be included in the QPRF report. Such compounds can be retrieved from the compound or dataset service using some similarity search algorithm. StructuralRelations are used to group together such analogues in the data model of the QPRF report (RDF). Here is an example:



Section 3.3.c

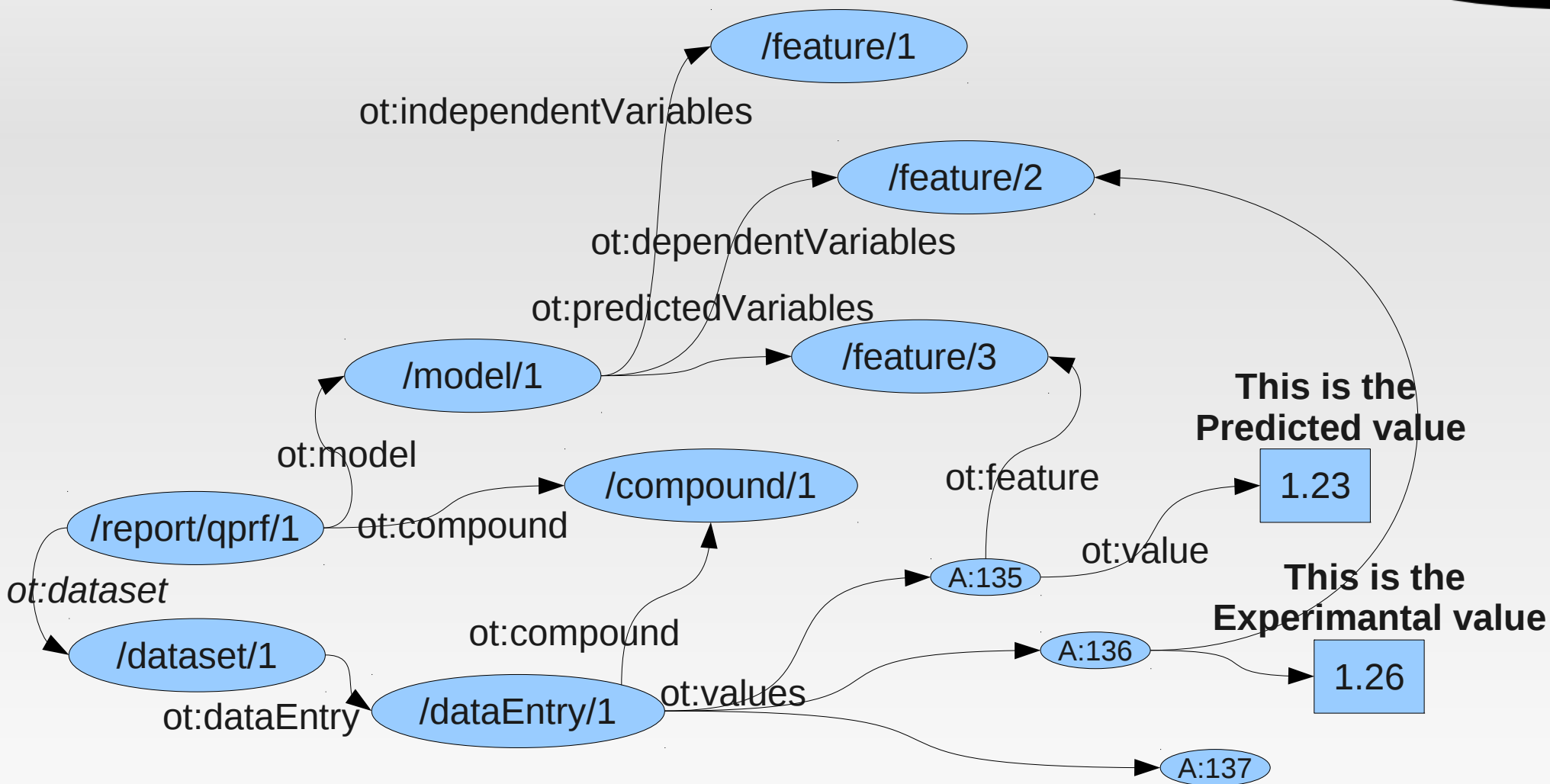
Compounds



Information that should be available for any compound contained in a QPRF report:

1. Compound Identifiers: {Smiles, InChI, InChI Key, CAS-RN, IUPAC Name}
2. Synonyms (Available chemical names in a colon separated list as a single string)
3. If available a link to an Image URI (otherwise the Image will be retrieved by the compound URI as an image/png or using some depiction service like AMBIT) – Introduction of the ontological class '*ot:Image*'
4. Data Entries (assigned to the QPRF report) containing all descriptor values for the compound that appear in the model (independent, dependent, predicted) – These are wrapped in a dataset assigned to the QPRF report.

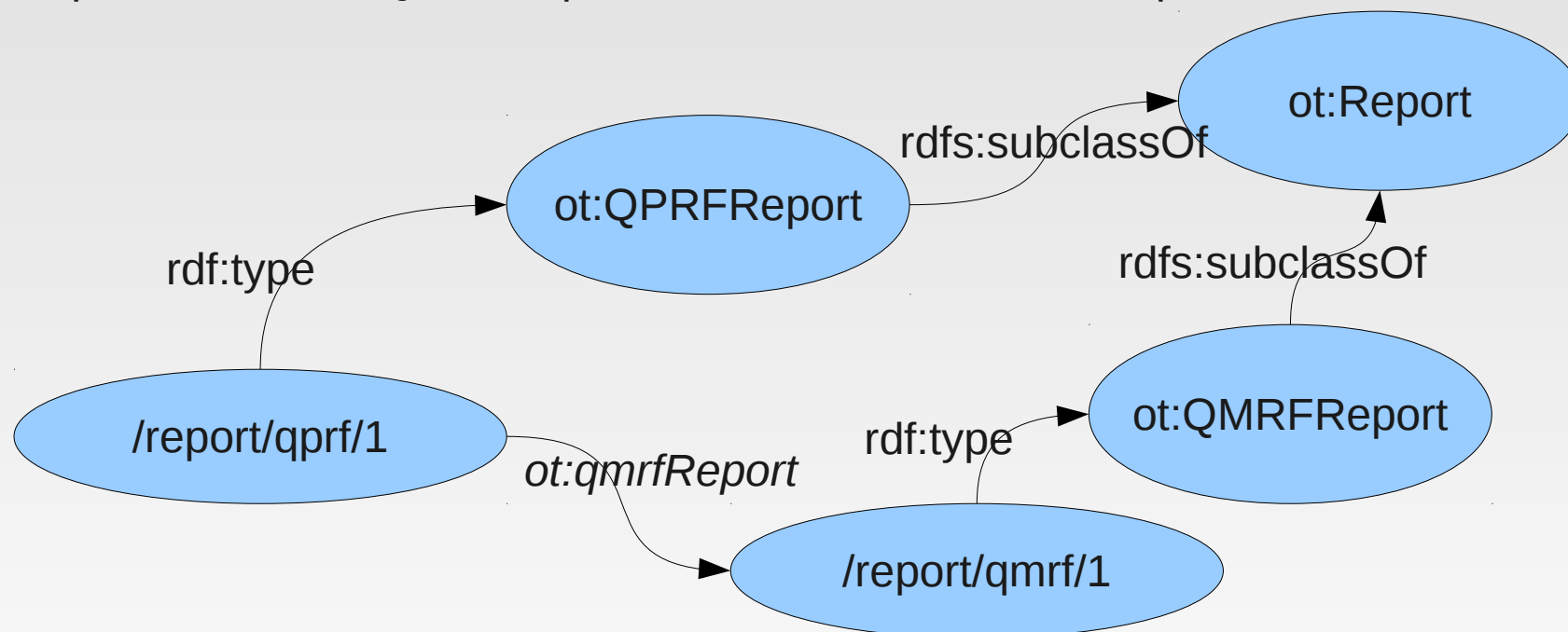
Descriptor Values - Example



Reference to QMRF Report



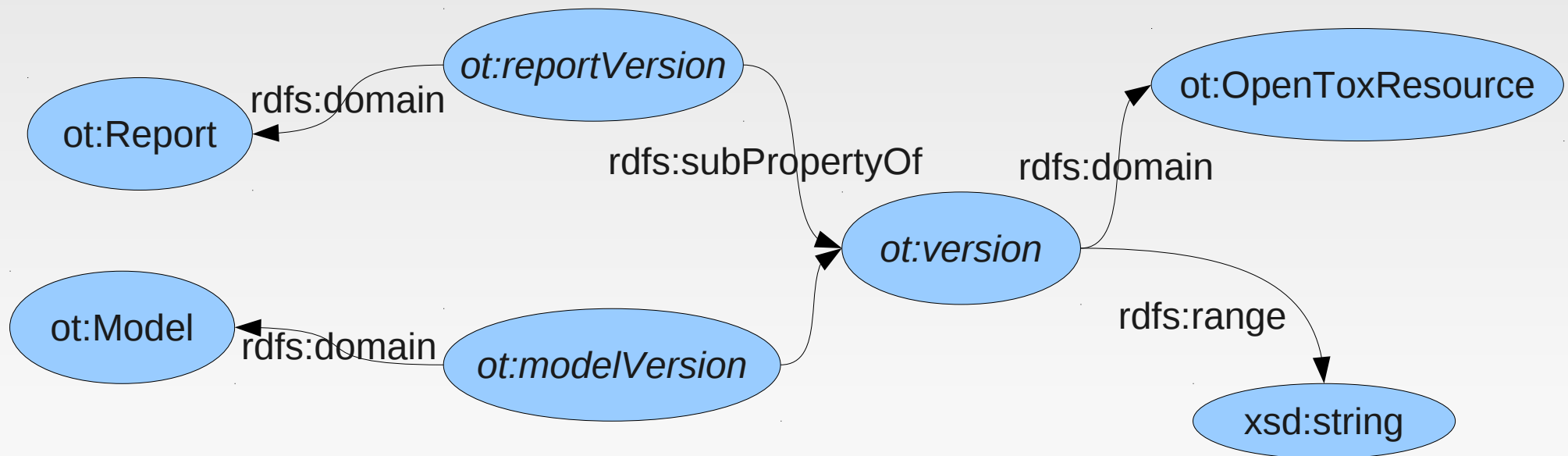
According to JRC specifications all QPRF reports should have (if possible) a reference/link to a QMRF report for the model that produced the prediction. This should be established by the QMRF reporting services (QMRF report URIs). In the RDF of a QPRF report, a link to a QMRF report can be a Node. For example:



Model and Report Versions



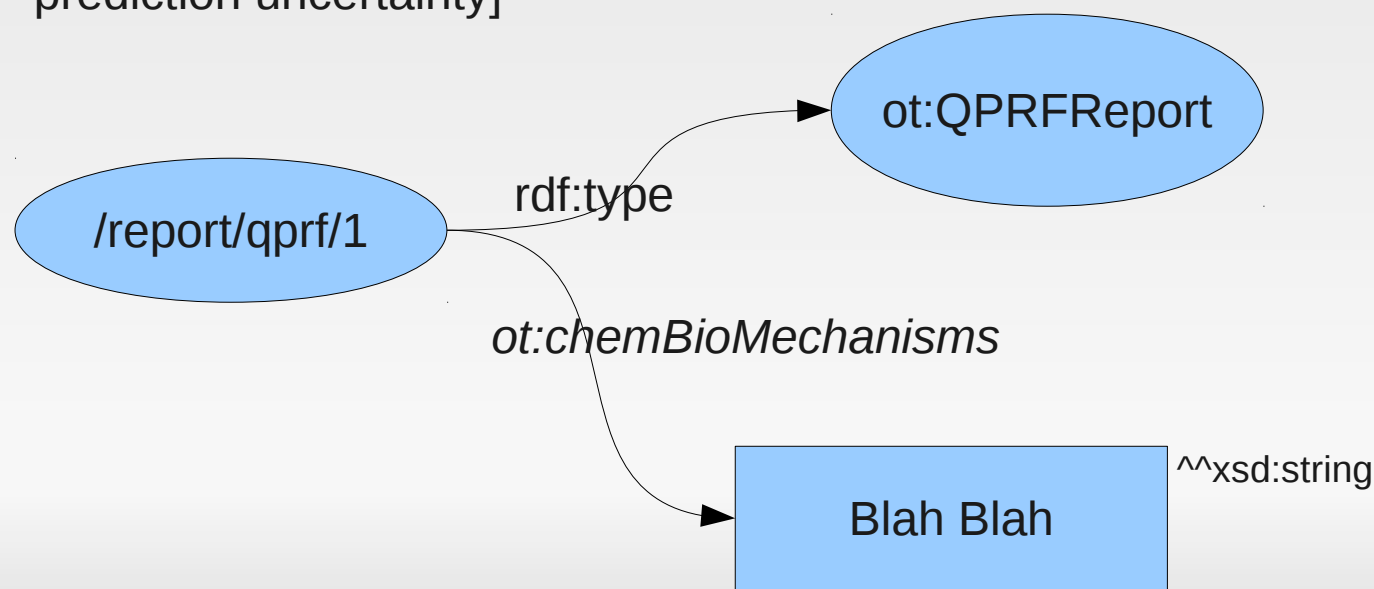
- Why not using owl:versionInfo: According to W3C OWL specifications 'the object of owl:versionInfo is a literal and the tag can be used to annotate **classes** and **properties** in addition to ontologies.'
- Use dc:date and ot:modelVersion and ot:reportVersion instead.



Prediction Value



The predicted value and (if available) the experimental value are assigned to the QPRF report base node as data entries (wrapped in a dataset). Each data entry is related to a compound (the compound for which the report is made) and a feature (in our case, either the dependent feature of the model [for the experimental value] or the predicted feature [for the predicted value – output of the model]). The comment appearing in section [3.2.e](#) of QPRF report can be assigned to the feature value as an **rdfs:comment** while we need additional properties to cater for the entries of [3.4.](#) and [3.5.](#) The latter can be assigned to QPRF report using the property *ot:chemBioMechanisms*. [See next slide for prediction uncertainty]



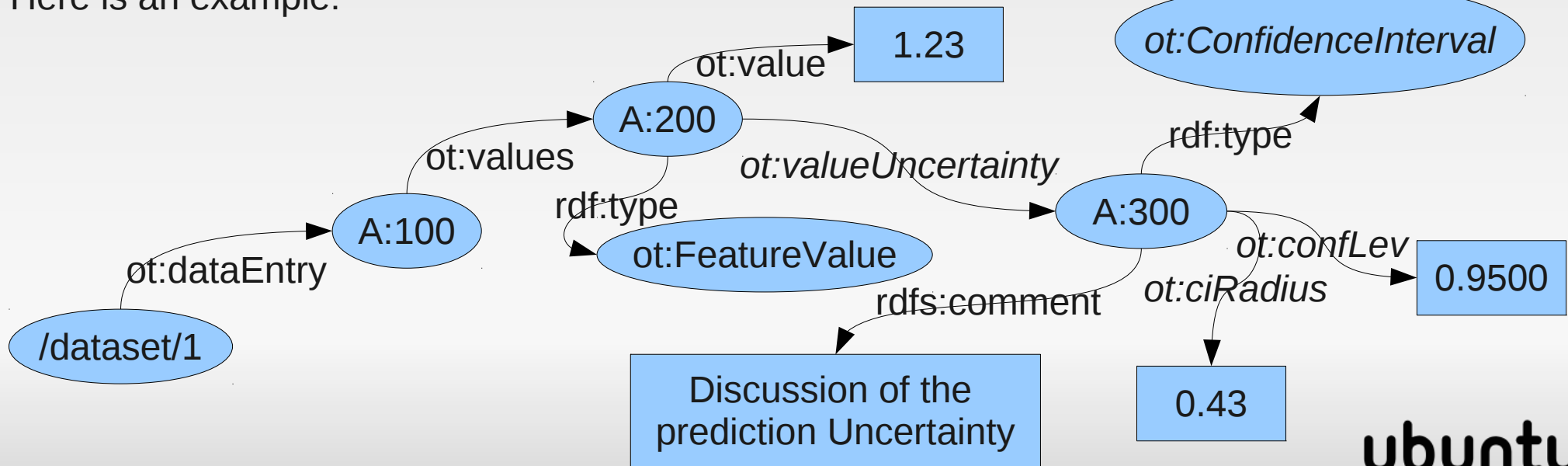
Prediction Uncertainty



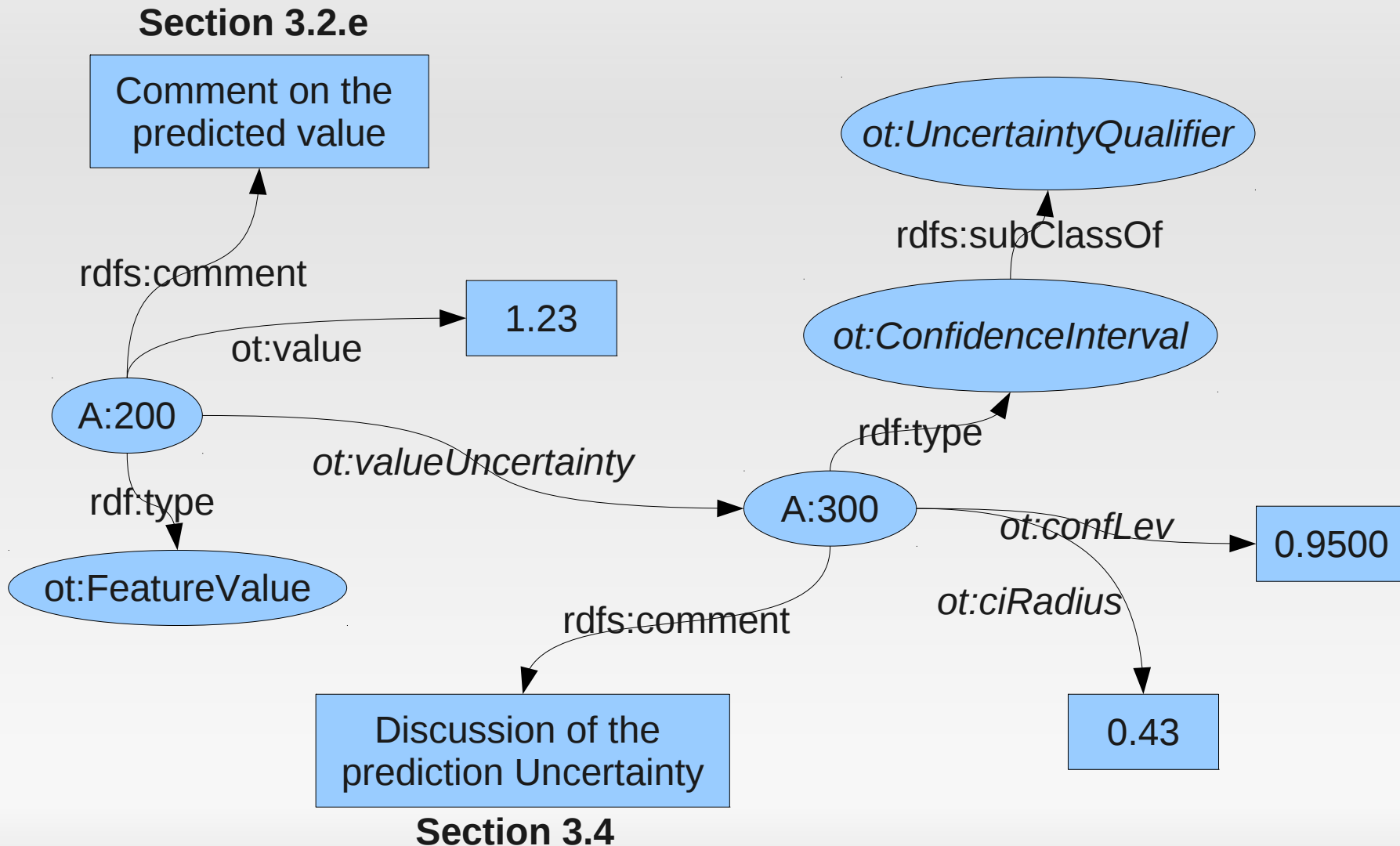
Taking into account that every value can be uncertain, we need a mechanism to allow us describe this uncertainty in any case. The uncertainty can be expressed using:

- * Error Bounds e.g. 10.0 ± 0.1
- * Confidence Intervals e.g. (0.9,10.1) using 95% confidence level.
- * Probability Values (for nominal values only)
- * Parametrized Distribution (e.g. $N(\mu, \sigma)$)
- * Significant Digits (straightforward)

To every node of type 'ot:FeatureValuePair' we can assign such information about the uncertainty of the value related to this node. Here is an example:



Sections 3.2.e & 3.4

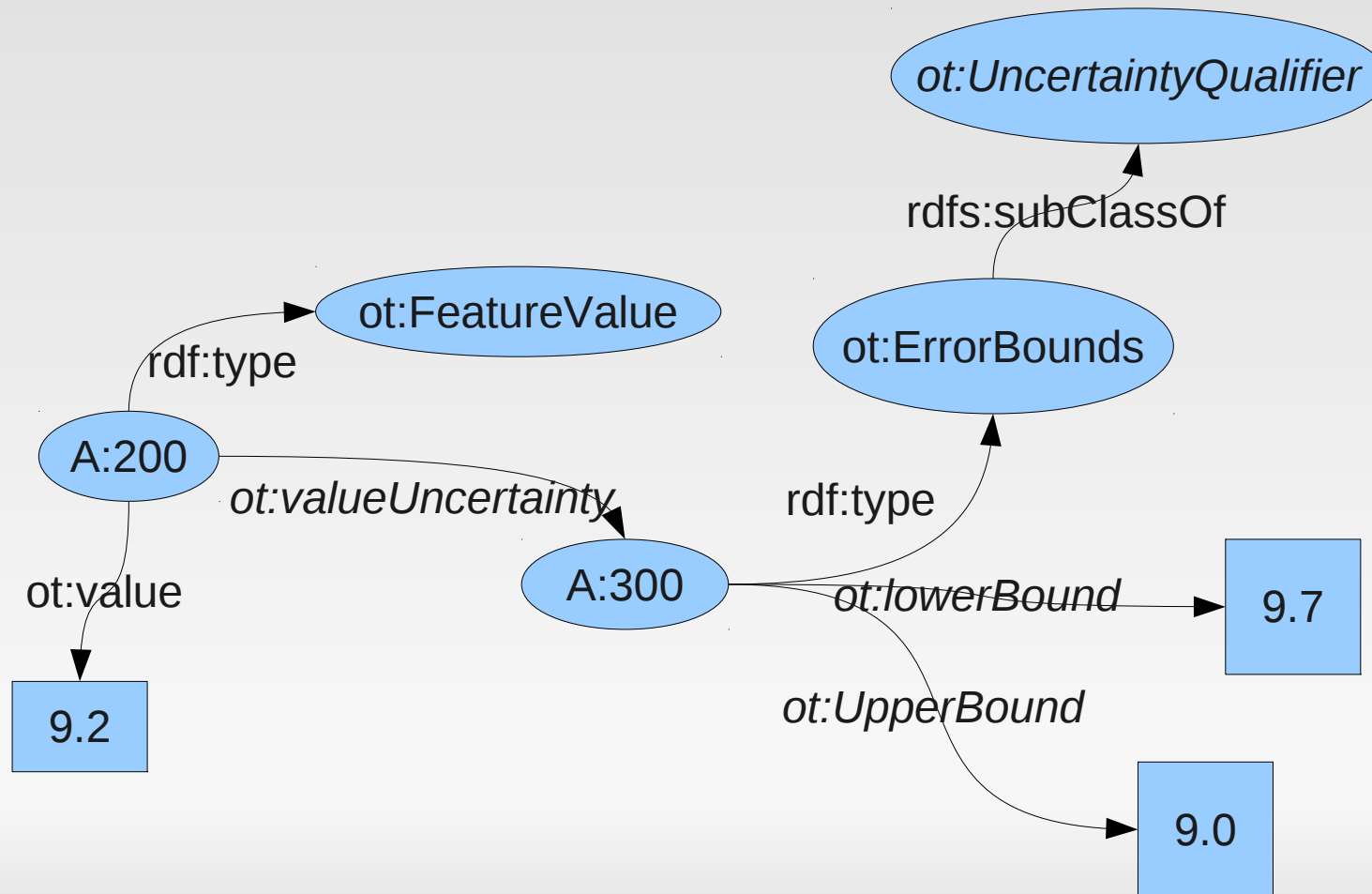


More About Uncertainty

Error Bounds



Here is an outline of the classes and properties needed to describe the uncertainty of a values in terms of Error Bounds:



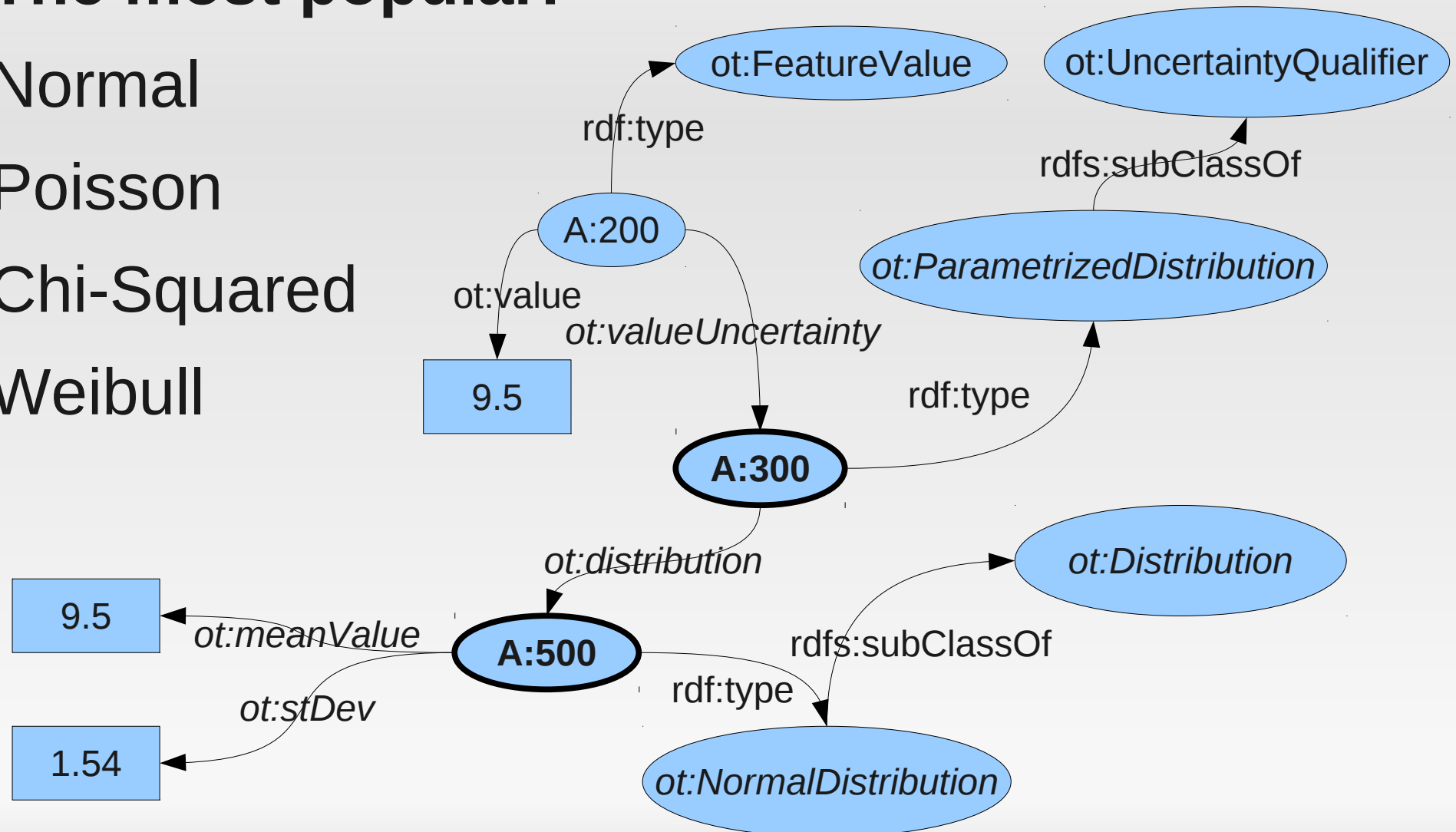
More About Uncertainty

Parametrized Distributions



The most popular:

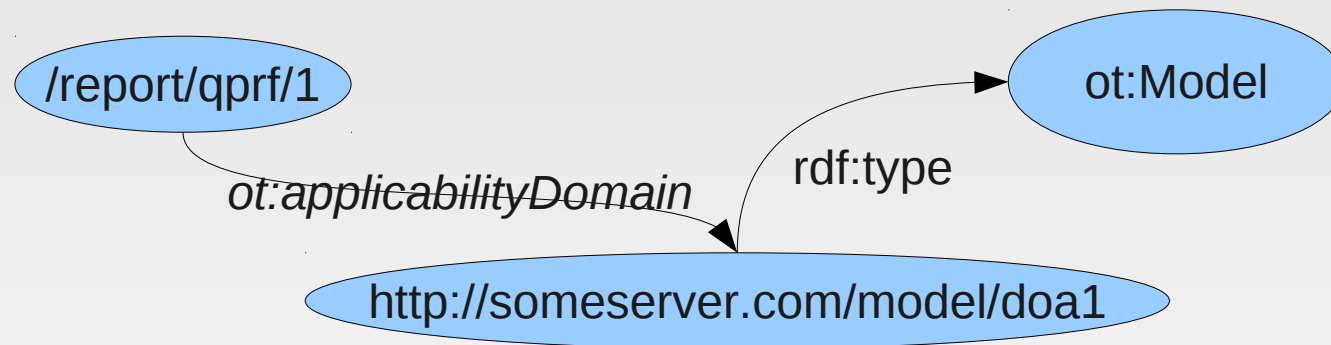
- Normal
- Poisson
- Chi-Squared
- Weibull



Applicability Domain

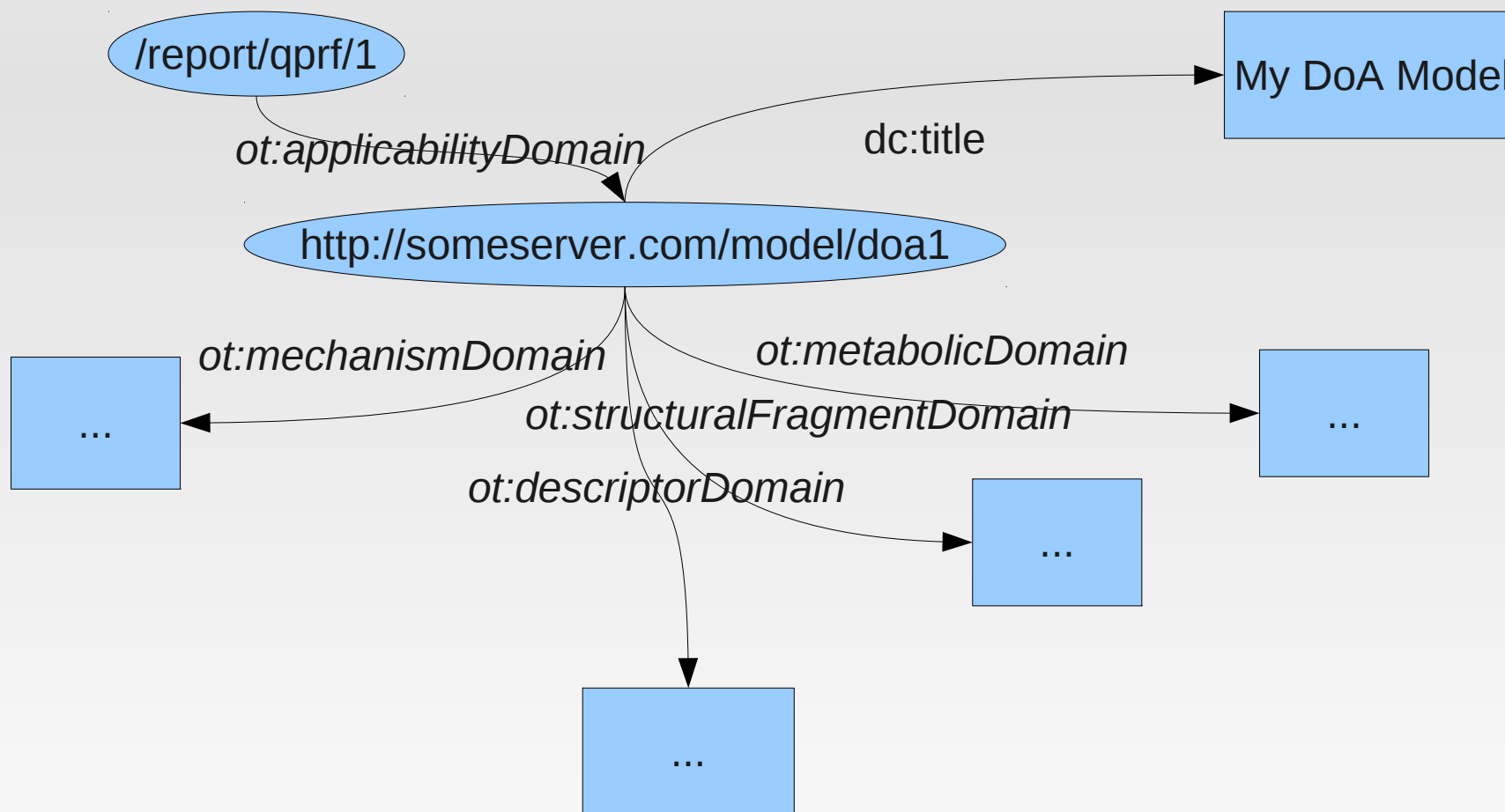


There should be an explicit reference to a DoA Model (if available). This should be provided by the client (e.g. the application that starts the editor can optionally provide a link to a DoA model resource.). If no such link is provided and there is a 'recommendation' in the model resource about a DoA training algorithm, then a DoA model can be (re)produced out of the training dataset of the model.



Meta information about the DoA model like the ones appearing in sections 3.3 and 3.3.a. can be assigned to this node.

DoA Additional Information



Important ToDos



- Discuss on this proposal; Input from developers
- Update OpenTox OWL file; Make it available on the site
- Make an example RDF file describing a QPRF report
- Extend Q-Edit to be able to create/parse such RDFs. Support Save/Open Functionality.
- Development of Reporting WS
- Holidays! (This is the hardest one but we can take that on)