

## **Qurator: An Ontology-Based Approach to Handling Information Quality in e-Science**

Paolo Missier, Suzanne Embury, Mark Greenwood (University of Manchester) Alun Preece, Binling Jin (University of Aberdeen)



Qurator seeks to assist scientists and data curators in managing the quality of their information. Rather than trying to impose a common set of generic IQ priorities on all users of a resource, an alternative is to provide scientists with the means of expressing explicit descriptions of quality that are relevant to their domain of interest and specific to their current task

Working closely with user-scientists in two post-genomics domains proteomics and transcriptomics - two hypotheses will be tested:

- 1. that it is possible to elicit detailed specifications of the IQ requirements of user-scientists, preferably in a formal language so that the definitions are machine-manipulable
- 2. that the annotation of information resources with detailed descriptions of their quality can be performed in a cost-effective manner



In transcriptomics, microarray experiment data is routinely captured in MAGE-ML format. Elements of an experiment should be described in a standard way using terms from the **MGED Ontology**<sup>\*</sup>.

In searching for microarray experiment data to use for their own purposes, a particular biologist may specify a quality requirement on the extent to which particular elements of the dataset – called **ontology entries** – conform to the MGED Ontology.



Contact info@qurator.org



## An ontology-based approach to IQ Ontologies play several roles in C myGrid informatics\_physical\_structure defining the Ourator IO T C myCrist data V C QualityMetric framework. The main concepts of Array test report the framework are defined in our C Impact.Factor.test.report **IO Ontology** (represented using C MOED, compliance, report MANE-complance-report the Web Ontology Language. C OntologyEntry-validation-resul OWL) which is aligned with the OPL myGrid project data ontology. While the framework allows for the definition of highly domain-specific (and scientist-specific) IQ preferences, it allows for the classification of these preferences under a generic categorisation drawn from the IO literature. **Description logic-style** We are experimenting with the IQ preference:

various possibilities afforded by the OWL-based representation for defining preferences in a machine-manipulable way.

Rule-style IQ preference: "a dataset is acceptable if more than 80% of ontology

"a dataset is acceptable if more than 80% of ontology entries fully comply and the experiment was performed in the past 2 years"

## Example IQ service: testing ontology conformance

As a concrete example of the Qurator approach, we have implemented an ontology-conformance testing Web service. The service requires two pieces of information:

- the URI of an XML document containing experiment data
- an XML control file specifying the elements to check in the experiment data (as XPath expressions)

The service returns a report detailing the conformance of each specified element. This conformance report can then be used to generate preference classifications and results for presentation.

## Information Quality Ontology Conformance

Please Choose the data (XML) file that you want to validate: C:\Documents and Settings\apresce\Desider(\_Browse...

Currently, we use a default control file. To view the default control file, click here

Valdate Reset



According to the delatik preference: the monitation quarky or your opcaded Unacceptable

Acceptable	=		VAL_OK	3	٧	80 M	9
	AND	٧	VAL_BAD_IND	¢#	٧	10 💌	9
	AND	۷	VAL_BAD_CLASS	<b>C</b> #	٧	5 v	9
Calculate	Rese						

The Web service is designed to handle a variety of different kinds of ontology entries. The MGED Ontology handler for the MAGE-ML experiment data uses the DAML version of the ontology, and is able to check conformance of both ontology **classes** and **individuals**. Other forms of ontology and controlled vocabulary in common use can also be checked with alternative handlers, including simple textual/lexical lists of terms, RDF Schemas, and the various forms of OWL.

Part of the aims of Qurator are to embed the IQ-management tools within the scientists' working environment. To this end we are currently creating alternative clients for the ontology conformance Web service, including a general-purpose Webbased interface (shown here), and a client plugin for the Pedro data entry tool widely used by biologists.

Evaluation Result	
URL MAD INC	(MAGE PELT) Protocol, packager's UProtocol, availed (1) Protocol 91/7
URL SAD INC	(NACE 18.11) Territoria package [] Compound market 1] Compose
	(MACE MULT) TRATING package(1) The Material annual (1) The Source
UNL BAD IND	ANCE AUTO TRANSPORT processor (1) The Hadenia and media Unicidentia
UNL BAD IND	(NACE PELIS) TO PARTY IS package (1) To Platence, and Ref (1) The four to
N. 540, NO	NOCE PELLINGHAMMENT DECKNOLULING AND
INL DAD, INC	Made multiliteratives package(1) to Material assess(1) to fourier
AL MO NO	(MACE HE_1) TO MAKE AN ANY 11 TO MAKE AN ANY 11 TO SOLVE
N, MO NO	MAGE-RELET TRANSPORTED and the Material and Millimscores
UN, DAD , NO	MAGE-MUTITE Meterial package[1]/McMaterial_accelet[1]/McGovern
NAL BAD JHD	MACE AN I I WORKSHIP DOCKSOFT I TO MATERIA AND AND I TO KNOW D
UK_BAD_BO	(MACE HE_II) TROMANNIE_DOCKOR(I) TROMANNE_MERINE(I) TROPOND
INC. SAD, INC	/MAGE ME/11/WoMsterne_package(1)/WoMsterne_neurist[1]/Womource
IN. SAD BO	(MAGE 46.11) ThoMake or package (1) the Material available (1) the Source
IN, SAO , NO	/NACE PERIOD TO PERFOR DOCKSOFT (Contract and Contract Statistics)
14,040,00	/94GE PE/11/Roberts_peckauf11/Roberts_antht11/Roberts
IN_MO_NO	/MAGE PRUID TRANSPORTED DOCKARED LINE PARTY AND INCOMENDATION OF THE PROPERTY AND INCOMENTATION OF THE PROPERTY AND INTERPOPERTY AND INTERPOPE
UNL_SHO_DID	MAGE MU[1] REPORTED_DOCKADE[1] RePorteD_MILITEDCOTTS
NE_SAD_BO	(MAGE PR.(1) The Processing processory 1) The Materian provided 1) The Source
NU.390,390	MAGE PR/11/TerManue_peckaur/11/terMateria_exemit/11/TerSource
NAL BAD BID	(MAGE 46.(1) Experiment_package(1) Experiment_assembl(1)/Experim



Visit www.qurator.org





0





The Qurator project is funded by the EPSRC Programme Fundamental Computer Science for e-Science: GR/S67593 & GR/S67609 - Describing the Quality of Curated e-Science Information Resources.