Context, Quality and Relevance: Dependencies and Impacts on RESTful Web Services Design

Hong-Linh Truong[1], Schahram Dustdar[1] , Andrea Maurino[2] , Marco Comerio[2]

[1]Distributed Systems Group, Vienna University of Technology
[2]Department of Informatics, Systems and Communication
University of Milano – Bicocca

truong@infosys.tuwien.ac.at
http://www.infosys.tuwien.ac.at/Staff/truong

VIT L AB
DISTRIBUTED SYSTEMS GROUP

Outline

- Motivation

- Quality, context and relevance in data-intensive services

- Access to quality and context associated with data resources

- APIs for specifying and obtained data quality

- Open issues
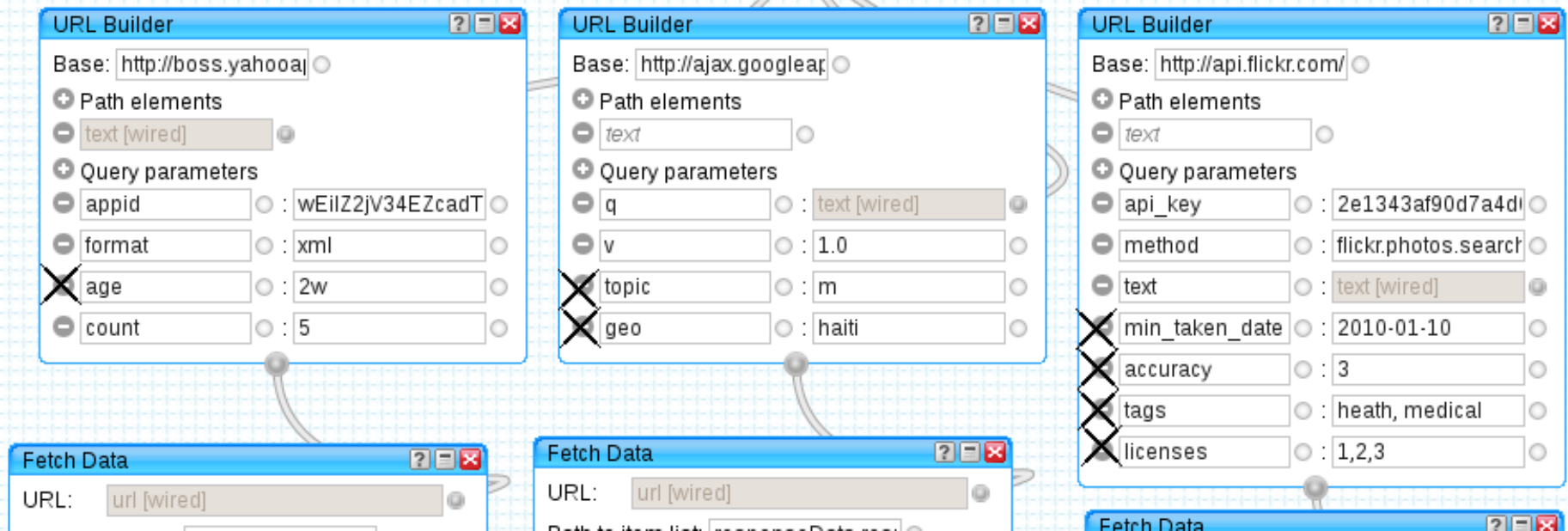
- Conclusions and future work

Motivation: compose data intensive RESTful services

- Composition of Yahoo! Boss News Search, Google News Search , and Flickr
  - recent news and high-qualified images, but free-of charge, related to "Haiti earthquake"
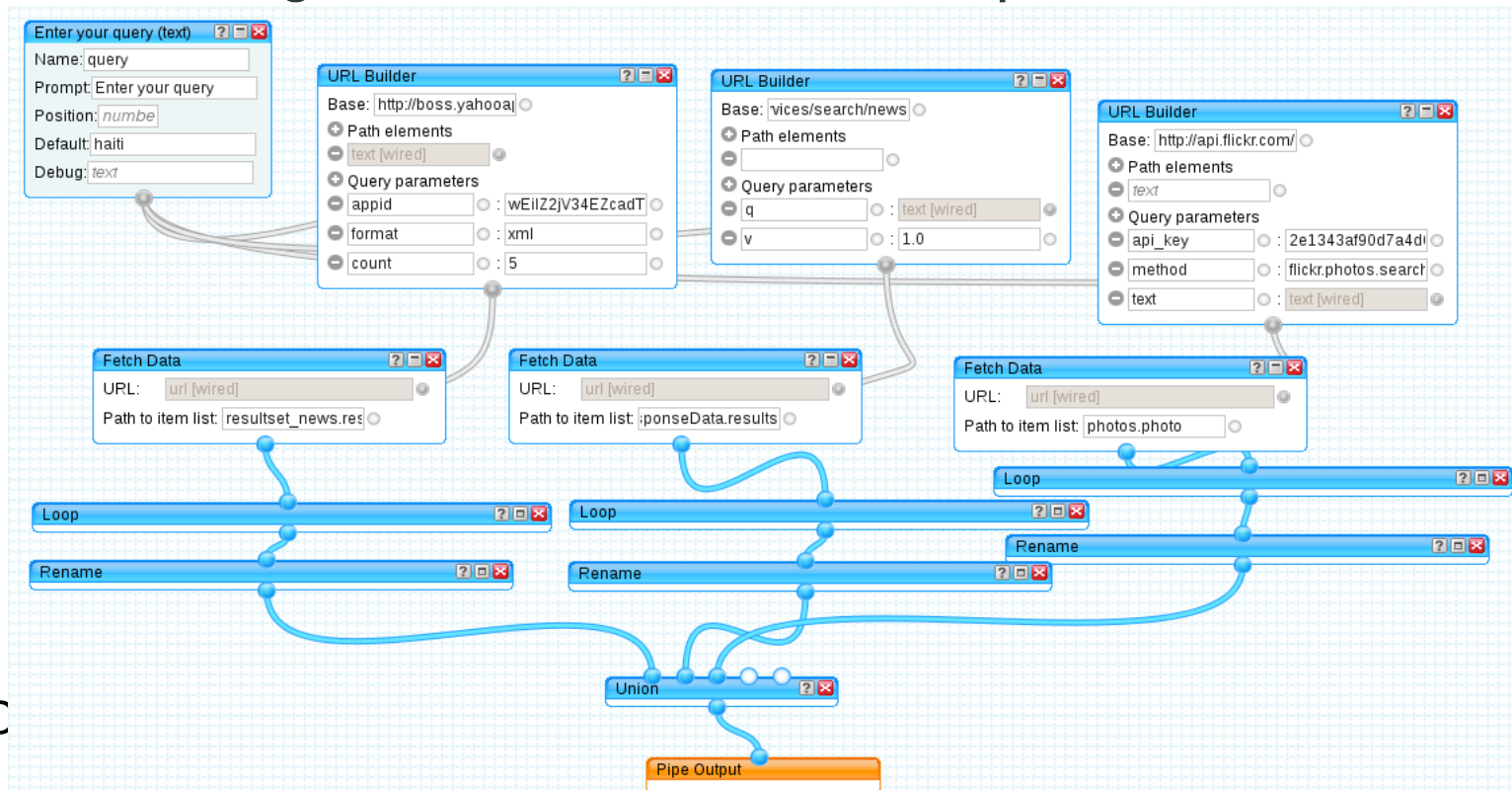
If the composer is aware of context and quality parameters

- Possible mappings of context and quality requirements

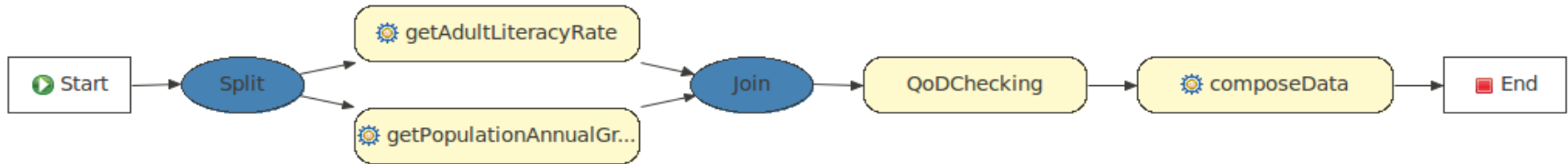but it is a tedious task and we are not sure we have a correct mapping

Motivation: access to large data resources

- Retrieve big datasets from RESTful services for further extraction, transform or mashup activities

```
−<datasets>
  <dataset organisation="IMF" category="International Financial statistics" name="BOP: capital account credit, US$"/>
  <dataset organisation="IMF" category="International Financial statistics" name="BOP: capital account debit, US$"/>
  <dataset organisation="IMF" category="International Financial statistics" name="BOP: current transfers credit, US$"/>
  <dataset organisation="IMF" category="International Financial statistics" name="BOP: current transfers debit, US$"/>
  <dataset organisation="IMF" category="International Financial statistics" name="BOP: exports of goods, f.o.b., US$"/>
  <dataset organisation="IMF" category="International Financial statistics" name="BOP: financial account, US$"/>
  <dataset organisation="IMF" category="International Financial statistics" name="BOP: imports of goods, f.o.b., US$"/>
  <dataset organisation="IMF" category="International Financial statistics" name="BOP: income credit, US$"/>
  <dataset organisation="IMF" category="International Financial statistics" name="BOP: income debit, US"/>
  <dataset organisation="IMF" category="International Financial statistics" name="BOP: net errors and omissions, US$"/>
  <dataset organisation="IMF" category="International Financial statistics" name="BOP: services credit, US$"/>
  <dataset organisation="IMF" category="International Financial statistics" name="BOP: services debit, US$"/>
  <dataset organisation="IMF" category="International Financial statistics" name="BOP: trade balance, goods and services, US$"/>
  <dataset organisation="IMF" category="International Financial statistics" name="BOP: trade balance, goods, US$"/>
  <dataset organisation="WHO" category="Demographic and socioeconomic statistics" name="Adolescent fertility rate"/>
  <dataset organisation="WHO" category="Demographic and socioeconomic statistics" name="Adult literacy rate (percent)"/>
  <dataset organisation="WHO" category="Demographic and socioeconomic statistics" name="Gross national income per capita (PPP international $)"/>
  <dataset organisation="WHO" category="Demographic and socioeconomic statistics" name="Net primary school enrolment ratio female (percent)"/>
  <dataset organisation="WHO" category="Demographic and socioeconomic statistics" name="Net primary school enrolment ratio male (percent)"/>
  <dataset organisation="WHO" category="Demographic and socioeconomic statistics" name="Population (in thousands) total"/>
  <dataset organisation="WHO" category="Demographic and socioeconomic statistics" name="Population annual growth rate (percent)"/>
  <dataset organisation="WHO" category="Demographic and socioeconomic statistics" name="Population in urban areas (percent)"/>
  <dataset organisation="WHO" category="Demographic and socioeconomic statistics" name="Population living below the poverty line (percent living on less than US$1 per day)"/>
  <dataset organisation="WHO" category="Demographic and socioeconomic statistics" name="Population median age (years)"/>
  <dataset organisation="WHO" category="Demographic and socioeconomic statistics" name="Population proportion over 60 (percent)"/>
  <dataset organisation="WHO" category="Demographic and socioeconomic statistics" name="Population proportion under 15 (percent)"/>
  <dataset organisation="WHO" category="Demographic and socioeconomic statistics" name="Registration coverage of births (percent)"/>
  <dataset organisation="WHO" category="Demographic and socioeconomic statistics" name="Registration coverage of deaths (percent)"/>
  <dataset organisation="WHO" category="Demographic and socioeconomic statistics" name="Total fertility rate (per woman)"/>
```

Done

http://www.undata-api.org/

ComposableWeb's 10, ICWE 2010, Vienna

5

VITALAB

VIENNA INTERNET TECHNOLOGIES ADVANCED RESEARCH LAB

DISTRIBUTED SYSTEMS GROUP

Motivation: access to large data resources

- Without QoD: get datasets and perform mashup
- With QoD support:
  - http://localhost:8080/UNDataService/data/query/Population annual growth rate (percent)?crq.qod:
    - dataelementcompleteness= 0.8654708520179372, datasetcompleteness=0.7356502242152466;
  - http://localhost:8080/UNDataService/data/query/Adult literacy rate (percent)?crq.qod:
    - dataelementcompleteness=0.5874439461883408,datasetcompleteness=**0.04349775784753363**
 - → *Should we retrieve the data and perform data mashup?*

Context, quality and relevance for data-intensive services

- Data-intensive services provide data resources whose quality and context could not easily described at *the level of service as a whole*

- We all know: missing context and quality leads to irrelevant results

- But several questions:
    - how can the composition developer recognize context and quality description associated with data resources?
    - how context and quality parameters can be mapped and passed to a service via REST APIs?
    - how to obtain context and quality description associated with data resources so that further activities can be made?

VIT/LAB
VIENNA INTERNET
TECHNOLOGIES
ADVANCED
RESEARCH LAB
DISTRIBUTED SYSTEMS GROUP

Related work

- Well-researched QoS and context WS
  - QoS-based and context-oriented Web service design
  - QoS-based/NFP-based Web service discovery and ranking
  - User experiences and context models for personalized Web services .
- However, existing works mainly deal with QoS and context information
  - QoD other data concerns are the key to data resources
  - Data resources have different data concerns
- Recent work on quality of mashup

Context, quality, and relevance dependencies in data service composition
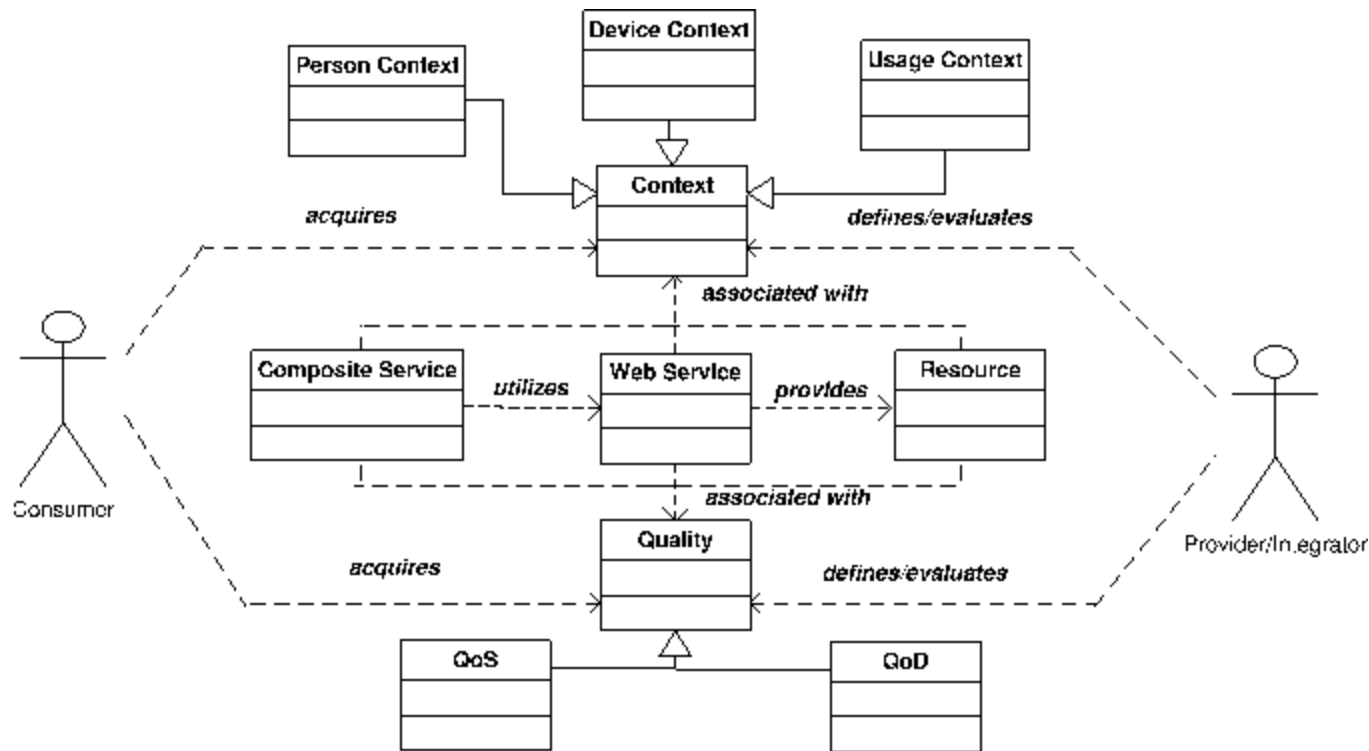
- Data resources: (semi-)structured data, unstructured documents, images, zipped datasets

Impact on the lack context and quality – at data resource level

- Difficult to select relevant resources and to provide *general description* of context and quality for the service managing the resource.

- The service has to build its own quality and context description

  - difficult if the service developer is not the data resource provider

- It is not clear if the resource will meet the consumer quality description

  - lead to the information overloading

  - challenge for the service to determine the quality of the data resource.

VIT/LAB
VIENNA INTERNET
TECHNOLOGIES
ADVANCED
RESEARCH LAB
DISTRIBUTED SYSTEMS GROUP

Impact on the lack context and data quality – at data service level

- Not sure if the service can be used.
- Not sure if resources provided can be used.
  - The composite service has to implement its own service selection mechanisms.
- The composite service is not sure about the quality of services and resources to be composed.
  - The composite service has to determine the quality and context of the resources at its side using its own knowledge.

VIT/\LAB VIENNA INTERNET TECHNOLOGIES ADVANCED RESEARCH LAB
DISTRIBUTED SYSTEMS GROUP

Access to data quality and context associated with data resources

- Data quality and context associated with data resources:
  - Different with representative quality and context associated with a service as a whole
- Current works:
  - Describe context and QoS associated with a service or a service operation but not with data resources
- We focus on mechanisms/convention to access quality and context associated with data resources
  - Not how to determine quality and context metrics

VIT LAB
DISTRIBUTED SYSTEMS GROUP

Data quality and context support in current RESTful WS design

- Many RESTful services do not provide data quality and context information

- In RESTful service description, WADL or ATOM
  - Allow specify **individual resource** but there is no description for context and quality information.

- At the REST APIs level, there is **no convention** for specifying and obtaining context and quality information.

  → provide context and quality description exchange protocols among the interactions of service compositions, services and resources

VIT∧LAB
VIENNA INTERNET TECHNOLOGIES ADVANCED RESEARCH LAB
DISTRIBUTED SYSTEMS GROUP

Data quality and context support in current RESTful WS design

- RESTful WS design: PUT, GET, POST, DELETE data resources but no support for non-functional properties (e.g., context/quality)

- Possible option:
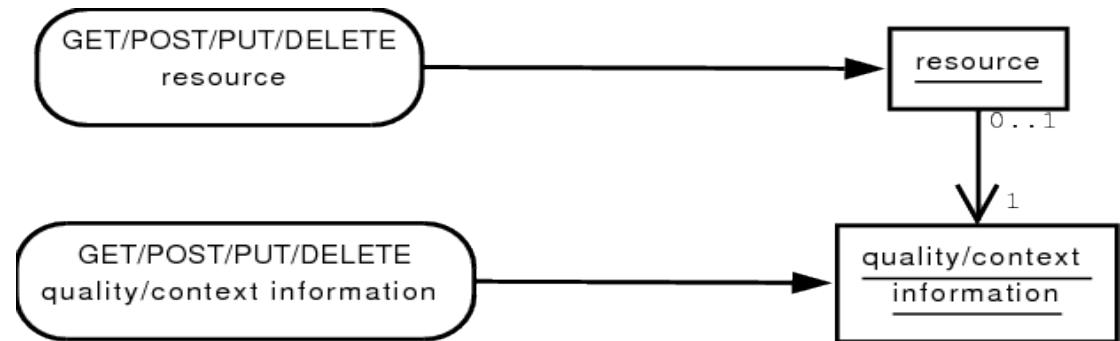
→ difficult to understand and manage the relationship between the resource and its quality/context description

- We assume: non-functional properties are a part of data resources → using the same operation for the same resource id but with different parameters

Proposed design principles – convention

- Focus on specifying and obtaining context and quality **associated with data resources** provided by services

- Static and dynamic context and quality binding
    - Static: context and quality can be published with resource description (e.g., WADL, ATOM)
    - Dynamic: context and quality information obtained at runtime

- We test this principle with
    - WADL, a RESTful service provides the Google Flu Trend data, a RESTful service wrapping the UN datasets service

VIT∧LAB
DISTRIBUTED SYSTEMS GROUP

Representations for Context and Quality

- We do not develop/propose new context and quality models
  - Assume that the community will rely on existing models
  - We use custom models from [APSCC 2009: 87-94]

```
{
 ...
  "crq.qod": {
     "crq.uptodateness": "up to dateness",
     ...
     "crq.dataelementcompleteness":"data element completeness",
     "crq.datasetcompleteness":"data set completeness",
     "crq.domainspecificqod": [
      {
                   "name": "name of the metric",
                   "value": "value of the metric"
      },
     ]
  },
 …
}
```

Static quality and context information for data resources WADL

specify context and quality schemas and their types using **grammars/include** and **doc/title**

```
<grammars>
    <include
href="http://www.infosys.tuwien.ac.at/prototype/SOD1/restfuldesign/schemas/crq-
quality.xsd">
        <doc title="Quality">
            grammar describing quality
        </doc>
    </include>
    <include
href="http://www.infosys.tuwien.ac.at/prototype/SOD1/restfuldesign/schemas/crq-
context.xsd">
        <doc title="Context">
            grammar describing context
        </doc>
    </include>
</grammars>
```

17

VIT LAB
VIENNA INTERNET
TECHNOLOGIES
ADVANCED
RESEARCH LAB
DISTRIBUTED SYSTEMS GROUP

Static quality and context information for data resources with WADL

specify the published, static quality and context information using **representation**

```
<representation id="QualityPresentation" element="crqQuality:crq.qod"
                    mediaType="application/json">
  <doc title="QoDDescription">
    {"crq.qod": {
            "crq.datasetcompleteness": 0.10256410256410256,
            "crq.consistency": 1,
            "crq.uptodateness":5
    }}
  </doc>
</representation>
<representation id="ContextPresentation" element="crqContext:crq.context"
                    mediaType="application/json">
  <doc title="ContextDescription">
    {"crq.context": {
            "crq.location": "Europe",
            "crq.datapermission":"free"
    }}
  </doc>
</representation>
```

Static quality and context information for data resource discovery with ATOM-alike model

```xml
<cdc:category label="Demographic and socioeconomic statistics"/>
  <cdc:id>http://undata-api.appspot.com/data/query/Adult literacy rate (percent)
</cdc:id>
  <cdc:title>UN Data on Adult literacy rate</cdc:title>
  <cdc:entry>
    <cdc:category term="DaaSConcerns" scheme="...daasconcern-v0.2.xsd"/>
    <cdc:content type="application/xml">
<tns:qod xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns:tns="http://www.infosys.tuwien.ac.at/SOD1/daasconcerns/custom" xsi:type="tns:QoD">
  <tns:crq.dataelementcompleteness>0.5874439461883408</tns:crq.dataelementcompleteness>
  <tns:crq.datasetcompleteness>0.04349775784753363</tns:crq.datasetcompleteness>
</tns:qod>
</cdc:content>
    <cdc:published>2010-06-23T22:23:30.557Z</cdc:published>
  </cdc:entry>
```

VITALAB
VIENNA INTERNET
TECHNOLOGIES
ADVANCED
RESEARCH LAB
DISTRIBUTED SYSTEMS GROUP

Specifying and obtaining data context and quality using parameters in REST APIs

- Specifying requests by using utilizing query parameters the form of crq.metricName=value

> GET/resource?crq.accuracy="0.5"&crq.location=''Europe"

- Obtaining contex and quality by using context and quality parameters without specifying value conditions

```
curl http://localhost:8080/UNDataService/data/query/Population annual growth rate
(percent)?crq.qod
{"crq.qod" : {
"crq.dataelementcompleteness ": 0.8654708520179372,
"crq.datasetcompleteness": 0.7356502242152466,
...
}}
```

- Context and quality schema can be also accessed via parameters

Open issues

- Data context and quality of data resources is a part of the contract/licensing of the data resource?

- The use of parameters can be achieved but
  - Context and quality metrics and representations are diverse → how do we compare quality/context metrics defined by different services

- We propose only convention/principle. To convince its use we need:
  - Use accepted terms, evaluate via real applications
  - Perform possible different mappings

- Data concerns must more than data quality and context
  - e.g., data privacy

VITALAB
VIENNA INTERNET
TECHNOLOGIES
ADVANCED
RESEARCH LAB
DISTRIBUTED SYSTEMS GROUP

Conclusions and future work

- Analyzed the importance of data context and quality support in RESTful data-intensive services

- Limitations of current RESTful Web service design with the respect to data context, quality and relevance

- Propose conventions to specify and obtain context and quality associated with data resources

- Future plans
  - Examine possible mappings and focus on validation
  - Composition of context and quality of data resources
  - Link data concern evaluation to data concern publishing for data services

Thanks for your attention!

Hong-Linh Truong

Distributed Systems Group
Vienna University of Technology

Austria

truong@infosys.tuwien.ac.at
http://www.infosys.tuwien.ac.at