EU Project No:601043 (Integrated Project (IP))

DIACHRON

Managing the Evolution and Preservation of the Data Web DIACHRON

Dissemination level:	
Type of Document:	
Contractual date of delivery:	
Actual Date of Delivery:	
Deliverable Number:	
Deliverable Name:	
Deliverable Leader:	
Work package(s):	
Status & version:	
Number of pages:	
WP contributing to the deliverable:	
WP / Task responsible:	
Coordinator (name / contact):	
Other Contributors:	
EC Project Officer:	Federico Milani
Keywords:	
Abstract:	
Some meaningful abstract here.	

Document History			
Ver.	Date	Contributor(s)	Description
0.1	24.06.2014	Jeremy Debattista	Created TOC and LaTeX Setup

TABLE OF CONTENTS

TABLE OF FIGURES

LIST OF TABLES

1 Introduction

1.1 Scope and Objectives

1.2 Context of this Document

1.3 Document Structure

2 Data Quality Framework

2.1 High-Level Architecture

2.2 Stream Processor

2.3 The Dataset Quality Ontology

2.4 Quality RESTful API Design

3 Libraries Used

4 Ranking Service

4.1 Data Quality Assessment Process

4.2 Data Quality Metrics

- Metric input is a quad ¡?s, ?p, ?o, ?g¿ -

4.2.1 Accessibility Category

4.2.2 Availability

Dereferenceability Metric
HTTP URIs should be dereferencable, i.e. HTTP clients can retrieve the resources identified by the URI. A typical web URI resource would return a `200 OK` code indicating that a request is successful and `4xx` or `5xx` if the request is unsuccessful. In Linked Data, a successful request should return a document (RDF) containing the description (triples) of the requested resource. In Linked Data, there are two possible ways which allow publishers make URIs dereferencable. These are the `303` URIs and the `hash` URIs[1]. Yang et. al [**?**] describes a mechanism to identify the dereferenceability process of linked data resource.

> *Calculates the number of valid redirects (303) or hashed links according to LOD Principles.*

This metric (listing **??**) will count the number of valid dereferenceable URI resources found in the subject (?s) and object (?o) position of a triple. The `isDereferenceable(resource)` method uses the rules defined in [**?**]. The metric will return a ratio of the number of dereferenced URIs (deref) against the total number of triples in a dataset (totalTriples). The expected range is [0..1], where 0 is the worst rating and 1 is the best rating.

[1] http://www.w3.org/TR/cooluris/

Algorithm 1 Dereferenceablity Algorithm

```
procedure INIT
    totalTriples = 0 ;
    deref = 0 ;
procedure DEREFERENCE($\langle ?s, ?p, ?o, ?g \rangle$)
    if (isURI(?s)) && (isDereferenceable(?s)) then deref++ ;
    if (isURI(?o)) && (isDereferenceable(?o)) then deref++ ;
    totalTriples++;
```

4.3 Visualisation of Quality Assessment

4.4 Ranking of Quality-Computed Datasets

5 Crawling Service

6 Conclusions