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

1 INTRODUCTION 4
1.1 SCOPE AND OBJECTIVES 4
1.2 CONTEXT OF THIS DOCUMENT 4
1.3 DOCUMENT STRUCTURE 4

2 DATA QUALITY FRAMEWORK 4
2.1 HIGH-LEVEL ARCHITECTURE 4
2.2 STREAM PROCESSOR 4
2.3 THE DATASET QUALITY ONTOLOGY 4
2.4 QUALITY RESTFUL API DESIGN 4

3 LIBRARIES USED 4

4 RANKING SERVICE 4
4.1 DATA QUALITY ASSESSMENT PROCESS 4
4.2 DATA QUALITY METRICS 4
4.2.1 ACCESSIBILITY CATEGORY 4
4.3 VISUALISATION OF QUALITY ASSESSMENT 4
4.4 RANKING OF QUALITY-COMPUTED DATASETS 4

5 CRAWLING SERVICE 4

6 CONCLUSIONS 4

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

4.2.1 Accessibility Category

4.3 Visualisation of Quality Assessment

4.4 Ranking of Quality-Computed Datasets

5 Crawling Service

6 Conclusions