COLLABORATIVE EUROPEAN DIGITAL ARCHIVE INFRASTRUCTURE

Project Acronym: CENDARI
Project Grant No.: 284432
Theme: FP7-INFRASTRUCTURES-2011-1
Project Start Date: 01 February 2012
Project End Date: 31 January 2016

Deliverable No. :	7.2
Title of Deliverable:	Data integration toolkit and repository (first stable release)
Date of Deliverable:	30.07.2014
Revision No.:	1.0 FINAL
WP No.:	7
Lead Beneficiary:	Matematicki Institut Sanu u Beogradu (MISANU)
Author (Name and email address):	Alexander Meyer (alexander.meyer@inria.fr) Alexander O'Connor (alex.oconnor@scss.tcd.ie) Bojan Marinkovic (bojanm@mi.sanu.ac.rs) Emiliano Degl'Innocenti (emiliano.degli.innocenti@gmail.com) Fabrizio Butini (fab.butini@gmail.com) Ivan Čukić (ivan@mi.sanu.ac.rs) Jun Zhang (jun.zhang@kcl.ac.uk) Laurent Romary (Laurent.Romary@inria.fr) Mark Hedges (mark.hedges@kcl.ac.uk) Milica Knezevic (knezevic.milica@gmail.com) Natasa Bulatovic (bulatovic@mpdl.mpg.de) Patrice Lopez (patrice.lopez@inria.fr) Wei Tai (wtai@scss.tcd.ie) Zdenek Uhlír (Zdenek.Uhlir@nkp.cz)
Dissemination Level:	PP = restricted to other programme participants

SEVENTH FRAMEWORK PROGRAMME

Nature of Deliverable:	R = report; D = demonstrator
Abstract:	The Data Integration Toolkit's (DIT) goal is to support the creation of semantic descriptions of the content of digital archives and mapping of the metadata from local archive formats into the metadata format. The toolkit includes Natural Language Processing (NLP) functionality for entity recognition, extraction and disambiguation within the content. A key component of the DIT is the repository, which is responsible for maintaining the serialized form of the research guides, archive descriptions, ontologies, metadata and other semantic material created or harvested by CENDARI. This document will discuss the sustainable infrastructure approach, and the principal software components, processes and metadata models for CENDARI. Areas of activity: ● Data acquisition ● NLP technologies ● Semantic description and interlinking of content ● Development of CENDARI Repository ● Integration of data from participating archives