Project no. 600663

PRELIDA

Preserving Linked Data
ICT-2011.4.3: Digital Preservation

D2.1 Establishment of the Working Group

Start Date of Project: 01 January 2013
Duration: 24 Months

Organisation Name of Lead Contractor for this Deliverable

Version 1.0

Document Information

Deliverable number: D2.1
Deliverable title: Establishment of the Working Group
Due date of deliverable: April 2013
Actual date of deliverable: April 2013
Author(s):   Grigoris Antoniou, Carlo Meghini, Anna Fensel, David Giaretta
Participant(s):   HUD, CNR, UIBK, APA
Workpackage: 2
Workpackage title: Organizational Support
Workpackage leader: CNR
Est. person months: 1
Dissemination Level: PU
Version: 1.0
Keywords: working group, selection criteria, analysis

History of Versions

Version	Date	Status	Author (Partner)	Description/Approval Level
0.1	19/4/13		Grigoris Antoniou (HUD)	First complete draft
1.0	26/4/13		Grigoris Antoniou (HUD)	Final version; review comments from CNR implemented

Abstract

To achieve its goals, the PRELIDA project utilizes a Working Group methodology. This document describes the basic duties of the Working Group, the selection criteria used, the confirmed members of the Group, an analysis of its composition, and identification of missing expertise that will be filled in advance of the Group's first meeting.

Table of Contents

1   Introduction

PRELIDA aims at building bridges across the Digital Preservation and Linked Data communities, with the view of:

(a) making the Linked Data community aware of the already existing outcomes of the Digital Preservation community; and

(b) working out challenges of preserving Linked Data that pose new research questions for the preservation community. These challenges are related to intrinsic features of Linked Data, including their structuring, interlinking, dynamicity and distribution.

More specifically, the objectives of PRELIDA are to:

1. Raise awareness of existing Digital Preservation solutions in the Linked Data community, and facilitate their uptake to provide support, where possible, for the long-term preservation of Linked Data.

2. Collect, organize and publish use cases related to the long-term access to Linked Data. The use cases will be contributed by stakeholders that have Linked Data in their long-term business strategies, or by third parties that see Linked Data as a market opportunity.

3. Create a comprehensive state of the art on the technologies related to Linked Data and Digital Preservation, and set up a Technology/Research observatory in order to identify the most significant actors in the area of solutions to Linked Data and Digital Preservation challenges.

4. Bring together internationally renowned scientists and representatives of key stakeholders from both communities, and highlight latest advances in their areas. Facilitate them in exchanging experiences and in discussing latest progress and findings in hard research problems relevant to their areas. Facilitate them in working out specific characteristics of Linked Data that make existing Digital Preservation solutions not fully satisfactory. Then, collectively identify relevant challenges and paths of addressing them in years to come.

5. Nurture the building of a multidisciplinary research, technology and application community around the domain of preserving Linked Data.

6. Build new strategic alliances among stakeholders in the Digital Preservation and Linked Data communities.

7. Draw attention of standardization bodies on the potential issues related to the preservation of Linked Data in their respective standards, and indicate paths for addressing them in years to come.

8. Influence the design of new research programs addressing the identified R&D priorities, bringing together Digital Preservation and Linked Data, in the 2020 agenda of the European Commission.

9. Largely disseminate the findings of the action all over Europe and beyond.

A crucial instrument towards achieving these aims and objectives is the establishment of a continuous Working Group, bringing together key researchers and stakeholders from the Digital Preservation and Linked Data communities. This document describes the main responsibilities of this Group, the selection criteria used, confirmed Group members, an analysis of the initial composition, including the identification of any gaps that need to be filled before the commencement of the Group's work, scheduled for June 2013.

1.1   Outline of document

- Section 2 describes the main responsibilities of the Working Group.

- Section 3 describes the selection criteria and key performance indicators, as defined in the Description of Work.
- Section 4 provides the list of confirmed members of the Working Group.
- Section 5 provides an analysis of the initial composition against the selection criteria and key performance indicators, and identifies any gaps that need to be filled.

2   Responsibilities of the Working Group

- To analyse the state of the art in Digital Preservation and Linked Data;
- To collect, organize and publish use cases related to the long-term access to Linked Data;
- To set up a Technology/Research observatory in order to identify the most significant actors in the area of solutions to Linked Data and Digital Preservation challenges;
- To identify the challenges in addressing the problem of preserving Linked Data;
- To deliver a R&D roadmap for addressing these challenges;
- To analyse the potential scientific, societal and economic impact of the identified research issues.

3   Selection Criteria

According to the Description of Work, the Working Group "will be composed of leading researchers and representatives of key sectors within the Digital Preservation and Linked Data communities. Overall, the group will consist of 25-30 experts, both from consortium members and external bodies. Care will be taken to: (a) Identify key sectors within the Digital Preservation and Linked Data communities, and ensure that appropriate representatives will participate in the Working Group; (b) Ensure a good balance between the Digital Preservation and Linked Data communities..."

A preliminary, non-exhaustive list of key stakeholders was already included in the DoW. This list served as the starting point in PRELIDA's effort to compose a strong, comprehensive and balanced Working Group.

Research centres / Universities:
- ATHENA-IPSYP, Greece
- DERI, Ireland
- FORTH, Greece
- Free University of Amsterdam, Netherlands
- Free University of Berlin, Germany
- Institut Josef Stefan, Slovenia
- TU Vienna, Austria
- University of Edinburgh, UK
- University of Leipzig, Germany
- Open University, UK

Industry / user groups:

- Organizations (e.g. national registries) adhering to the Open Government Partnership (currently 8 founding governments, and other 43 national governments committed to the idea of Open Government)
- Cities and regions opening up their data (e.g. Vienna in Austria, Cluj-Napoca in Romania, Tirol region in Austria)
- Encyclopaedia services (e.g. Wikipedia with its associated initiative to extract structured information from Wikipedia - DBpedia),
- Online newspapers (e.g. New York Times) and news agencies (e.g. BBC),
- Publisher industry (e.g. Kluwer, Elsevier) and indexing services (e.g. DBPL),
- Companies providing RDF triple stores such as Garlik (4store, UK), Systap LLC (BigData, USA), Franz Inc (AllegroGraph, USA), Clark&Parsia (Stardog, USA). StrixDB (France), Dydra (Dydra, USA), OpenLink (Virtuoso, UK), Ontotext (OWLIM, Bulgaria), as well as companies providing graph database stores such as Orient Technologies (OrientDB, Italy), InfiniteGraph (USA), Kobrix Software (HyperGraphDB, USA), Neo Technology (Neo4j, Sweden), Sparsity Technologies (DEX, Spain)
- Domain-specific data-sharing projects/initiatives (e.g. the Open Pharmacological Concepts Triple Store initiative in the pharmacological domain, the National Center for Biomedical Ontology in the biomedical domain), Linked Open Commerce in the area of e-Commerce, LinkedGeoData in the geospatial domain
- Scientific databases owners and users (e.g. GenBank, UniProt, SciFinder, Web of Science)
- Domain-specific data integration solution providers (e.g. solution providers such as Genomatix, GeneGo, NextBio, etc, in the biomedical domain providing warehousing of biomedical data sets)
- Geospatial data providers (e.g. national mapping agencies, geological surveys)
- Environmental modelling and forecasting communities (e.g. weather service providers)
- Museums (e.g. British Museum)
- Digital Public Library of America
- Library of Congress
- Open Planets Foundation
- Internet/web archiving, e.g. Internet Memory Foundation and Hanzo
- National Science Foundation
- Airbus
- IBM
- Microsoft
- UK National Archives (TNA)
- CERN, ESA, Helmholtz, STFC, CSC
- Libraries: British Library, DNB (Deutsche Nationalbibliothek, German National Library), KB (Koninklijke Bibliotheek, National Library of the Netherlands), ONB (Österreichische Nationalbibliothek, National Library of Austria)
- BIG: Big Data Public Private Forum, Coordination Action in the Information Society Technologies (IST) Programme for Research, Technology Development & Demonstration under the 7th Framework Programme

4 Members of the Working Group

This is the current composition of the Working Group. An asterisk indicates pending response; all other listed members are confirmed.

Consortium:
- Grigoris Antoniou, HUD
- Anna Fensel, UIBK
- David Giaretta, APA
- Antoine Isaac, Europeana
- Carlo Meghini, CNR

Research centres / Universities:
- Kevin Ashley*, DCC Director
- Soren Auer, Leipzig
- José Borbinha, INESC
- Peter Buneman, Edinburgh
- Peter Burnhill, Edinburgh
- Vassilis Christophides, FORTH
- Richard Cyganiak, DERI
- Fausto Giunchiglia, Trento
- Elena Simperl, Southampton
- Herbert van de Sompel, Los Alamos National Laboratory

Industry:
- Francois Bancillon*, Data Publica
- Michael Factor*, IBM Haifa
- Fabrizio Gagliardi, Microsoft
- Rob Sharpe*, TESSELLA
- Syaufeng Lyou*, Google

User groups:
- Reinhard Altenhoner, German National Library
- Juan Bicarregui*, STFC
- Mariella Guercio, Sapienza e Digilab Roma
- Jamie Shiers, CERN
- Matthew Woollard* , UKDA

International bodies:
- Phil Archer, W3C / EU OPEN DATA STRATEGY
- Jan Brase, DATACITE
- Norman Paskin, International DOI Foundation

5 Analysis of Initial Working Group Composition

- So far 21 members have confirmed their participation, while replies to 7 invitations are pending.
- The number of confirmed participants provides a good basis for the operation of the Working Group, but is still below the target of 25-30. The target will be reached if most of the pending responses are positive; otherwise alternatives will be identified and invited.
- There is a reasonable balance between the areas of Digital Preservation and Linked Data. The former is stronger in the groups of industry and user groups, whilst the latter is stronger in terms of academic members. The Consortium will seek to invite 1-2 additional academic members closely aligned to the area of Digital Preservation.
- Industry participation will be excellent if pending replies are positive, but is insufficient on the basis of confirmed participants. The Consortium will closely monitor this aspect, will follow up on pending answers, and will seek alternatives ahead of the Opening Workshop.