



Project no. 600663

PRELIDA

Preserving Linked Data
ICT-2011.4.3: Digital Preservation

D2.7 Report on the second Summer School

Start Date of Project: 01 January 2013

Duration: 24 Months

UNIVERSITY OF INNSBRUCK

Version: final

Project co-funded by the European Commission within the Seventh Framework programme

Document Information

Deliverable number: D2.7
Deliverable title: Report on the second summer school
Due date of deliverable: 10|2014
Actual date of deliverable: 10|2014
Author(s): José María García
Participant(s):
Workpackage: WP2
Workpackage title: Organizational Support
Workpackage leader: CNR ISTI
Est. person months: 4
Dissemination Level: PU (Public)
Version: final
Keywords:

History of Versions

Version	Date	Status	Author (Partner)	Description/Approval Level
1	2014/10/24		José María García (UIBK)	First draft
2	2014/10/31		José María García (UIBK)	Scientific outcome section updated

Executive Summary

This deliverable reports on the second summer school organized by the PRELIDA project. The school ran in conjunction with the ESWC Summer School 2014 from September, 1st to September 6th, 2014, in Kalamaki, Crete, Greece, and focused on various Linked Data and Digital Preservation topics. Experts from both communities were invited to discuss the current challenges in the area, focusing on specific aspects of preservation of linked data to give insights for prospective PhD students and establish new connections between scientists.



Table of Contents

Document Information	1
Executive Summary	1
1. Introduction	3
2. Participants	4
3. Programme of the Summer School	5
4. Scientific Outcome	8
5. On-line Resources	10
6. Conclusion.....	11
References	12



1. Introduction

The main objective of the PRELIDA project is to raise awareness of existing Data Preservation solutions in the Linked Data community and provide a comprehensive state-of-the-art in technologies related to Linked Data and Digital Preservation fields resulting in a multidisciplinary research and technology community around preserving linked data domain.

The scientific dissemination of the PRELIDA project mission, objectives and outcomes is achieved through a number of planned activities including the organization of two summer schools. The main objective of the summer school is to raise awareness of digital preservation of linked data in related communities, expose the challenges and opportunities to the students of the PRELIDA domain and engage new stakeholders. In this second edition of the summer school, a specific track were devoted to Linked Data preservation, comprising tutorials and hands-on sessions where PRELIDA outcomes were presented in a practical form to interested students attending also the ESWC Summer School. Thus, the background of the students were not only on Linked Data and semantic technologies, but also on digital preservation.

As part of the regular ESWC programme and the specific additions of the Linked Data Preservation track, senior experts from Linked Data and Digital Preservation communities, both from the Academia as well as from the Industry, were invited to give keynote talks and provide tutorial and hands-on sessions during the school. This year the summer school invited two key experts from the British Museum working on the ResearchSpace¹ platform. The platform is based on Linked Data, aimed at supporting collaborative search, sharing and publication on the Web in the field of cultural heritage scholarly community. Furthermore, among senior experts from the Academia, the summer school invited experts from Digital Archives and PRELIDA Working Group participants to present students with their vision and contributions to the preservation of Linked Data. In addition, professionals from big data, ontology management and other related fields contributed in highlighting the latest trends in supporting sectors and facilitated the interaction and discussion in these fields.

This deliverable reports on the 2nd PRELIDA Summer School held in conjunction with the ESWC 2014 Summer School on September 1-6, 2014 in Crete, Greece. The deliverable is structured as follows. Chapter 2 lists the participants of the summer school. In Chapter 3 the main programme is presented explaining major sessions of the school. Chapter 4 discusses the scientific outcome of the summer school w.r.t linked data preservation. In Chapter 5 information about online resources is provided. Finally, Chapter 6 summarize the deliverable providing some concluding remarks.

¹ <http://www.researchspace.org/>



2. Participants

The summer school included 36 students in total, as well as tutors, keynote speakers and an organizing team.

The tutorial team included the following experts:

Maribel Acosta (Karlsruhe Institute of Technology)
Isabelle Augenstein (University of Sheffield)
Sotirios Batsakis (University of Huddersfield)
Irimi Fundulaki (Institute of Computer Science-FORTH)
René van Horik (Data Archiving and Networked Services - DANS)
José María García (University of Innsbruck)
Barry Norton (The British Museum)
Dumitru Roman (SINTEF ICT)
Elena Simperl (University of Southampton)

In addition, several experts from the Academia were invited as keynote speakers to provide the state-of-the-art in the development of Linked Data and Digital Preservation fields and included the following speakers:

Abraham Bernstein (University of Zurich)
John Domingue (The Open University)
Marko Grobelnik (Jozef Stefan Institute)
Dominic Oldman (The British Museum)
Guus Schreiber (VU University Amsterdam)

Moreover, the organizing team was coordinating the running of the school and comprised:

Aneta Tumulowicz
Fabian Flöck

Finally, the PRELIDA track organization team comprised:

Grigoris Antoniou (University of Huddersfield)
José María García (University of Innsbruck)

3. Programme of the Summer School

The summer school lasted 6 days starting from Monday morning till Saturday midday and included the opening session, tutorials, keynote speeches, discussion slots, hands-on and poster sessions, and was concluded by students' project works.

- **An opening session** was aimed to welcome all the attendees, to introduce the tutorial team, outline the program of the school for the next days, shortly introduce the projects, which supported the organization of the summer school including the PRELIDA project.
- **Tutorials** provided advanced interactive activities between tutors and students in the form of presentations and discussions on various Semantic Web and Preservation topics such as:
 - o **Linked Data and Data Management (1) module** introduced the basic Linked Data principles, best practices, how the data could be published, queried and provided various use cases of it. In detail, there was a separate slot on querying Linked Data and its underlying technologies, i.e. SPARQL Protocol and RDF Querying Language. Furthermore, Linked Data provision was a special concern of the tutorial session to raise awareness in recent supporting technologies.
 - o **Ontologies (2) module** focused on presenting the best practices in ontology engineering: the most successful approaches and frameworks, examples and use cases.
 - o **Natural Language Processing (3) module** aimed at describing techniques for analysing texts and obtaining semantically meaningful information from them, introducing GATECloud, NLP and corresponding tools.
 - o **Social Semantic Web (4) module** presented collaborative technologies for the Semantic Web, including how to use semantic technologies to enable interoperability in social software and especially in crowdsourcing.
 - o **Linked Data Preservation (5) module** focused on raising awareness on the digital preservation of Linked Data in related communities, exposing the challenges and opportunities associated with this area, and introducing current solutions.
- **Hands-on sessions** aimed at solidifying knowledge acquired during the tutorials and applying above mentioned semantic Web technologies in several educational use cases. The students worked individually as well as in teams to solve certain assignments with various semantic Web tools.
- **Keynotes** were aimed to introduce the most vital research problems in semantic Web field with particular focus on Linked Data, Data Analytics and Digital Preservation.
- **Poster sessions** aimed to provide the possibility for each student to present the research progress to senior researchers as well as to other students. Initially, each student was asked to bring a poster on the research topic she/he worked with during her/his master, diploma, PhD studies or scientific project in which the student was involved. The reviewer from the tutorial team was assigned to each student to ensure proper scientific scholarly peer review.
- **Student project** had an objective to apply acquired knowledge during the tutorials and hands-on sessions and stimulate team work. Students were asked to form the groups of 4-5 participants and decide on the research question from presented topics they would like to work on and provide implementation (demo software) to solve that problem. Furthermore, the students were recommended to reuse tools, datasets, approaches described during tutorials and keynote sessions. The project was concluded by presentation of results conducted by a team representative in front of all participants.

- **Award ceremony and closing** was dedicated to give the feedback to students' projects and award the best solutions.

The detailed programme of the summer school is provided below:

Monday, September 1st

08:15	09:00		<i>Registration</i>
09:00	09:15	John Domingue	Opening session
09:15	10:15	John Domingue	Keynote: An Introduction to the Semantic Web
10:15	10:45		<i>Coffee break</i>
10:45	12:30	Barry Norton	Tutorial: Fundamentals of Linked Data: main standards and technology components, motivating application scenario
12:30	14:00		<i>Lunch</i>
14:00	15:00	Irini Fundulaki	Tutorial: Querying Linked Data: SPARQL 101
15:00	15:30		<i>Coffee break</i>
15:30	17:00	Barry Norton, Irini Fundulaki	Hands-on: Semantic Web languages and standards: RDF, RDFS, SPARQL
18:30	-	Elena Simperl	Poster session and opening reception, <i>dinner</i>

Tuesday, September 2nd

9:00	9:15		<i>Administrative issues</i>
9:15	10:15	Marko Grobelnik	Keynote: Global Media Monitoring
10:15	10:45		<i>Coffee break</i>
10:45	12:30	Elena Simperl	Tutorial: Building and using ontologies
12:30	14:00		<i>Lunch</i>
14:00	14:30	Fabian Flöck	Student project kick-off
14:30	15:30	Maribel Acosta	Tutorial: Providing and consuming Linked Data
15:30	16:00		<i>Coffee break</i>
16:00	17:00	Maribel Acosta	Hands-on: Publishing and consuming Linked Open Data
18:30	-	Elena Simperl	Poster session continued and <i>dinner</i>

Wednesday, September 3rd

09:00	09:15		<i>Administrative issues</i>
09:15	10:15	Dominic Oldman	Keynote: Conflicting Cultures of Knowledge
10:15	10:45		<i>Coffee break</i>
		Barry Norton	Tutorial: Linked Data for NLP
10:45	11:30	Dumitru Roman	Linked Data Preservation Track Tutorial: Data publication and hosting solutions

11:30	12:30	Barry Norton, Isabelle Augenstein	Hands-on: Using Linked Data and GATE
		René van Horik	Linked Data Preservation Track Hands-on: Data management
12:30	14:00	Elena Simperl	<i>Lunch</i>
14:00	15:00	Sotirios Batsakis	Tutorial: Social Semantic Web and crowdsourcing
		Maribel Acosta	Linked Data Preservation Track Tutorial: Challenges on preserving Linked Data
15:00	16:00	José María García	Hands-on: Using Mechanical Turk to solve Linked Data problems
16:00	-		Linked Data Preservation Track Hands-on: Preserving Linked Data: The DBpedia use case <i>Excursion and dinner</i>
Thursday, September 4th			
09:00	09:15		<i>Administrative issues</i>
09:15	10:15	Abraham Bernstein	Keynote: Processing Linked Data at Warp Speed
10:15	10:45		<i>Coffee break</i>
10:45	11:30	All tutors	Q & A Session
11:30	12:30		Student project work (tutors available)
12:30	14:00		<i>Lunch</i>
14:00	18:00		Student project work (tutors available)
20:00	-		<i>Social Dinner</i>
Friday, September 5th			
09:00	09:15		<i>Administrative issues</i>
09:15	10:15	Guus Schreiber	Keynote: Knowledge Engineering and the Web
10:15	12:30		Student project work (tutors available upon request)
12:30	14:00		<i>Lunch</i>
14:00	18:00		Student project work (tutors available upon request)
20:00	-		<i>Dinner and party</i>
Saturday, September 6th			
09:00	12:00		Student project presentations
12:00	12:30		<i>Break and deliberation of judges</i>
12:30	13:00		Award ceremony and closing

4. Scientific Outcome

During the Summer School, several sessions were related to Linked Data preservation, including a specific separate track devoted to showcasing current challenges and solutions to interested students of the summer school.

First, the keynote given by Dominic Oldman [1] discussed precisely how two traditionally separated communities, such as the scientific community and humanities community, can collaborate and take advantage of the other community's efforts. Specifically, he presented the ResearchSpace project been carried out by the British Museum, where Linked Data technologies are being applied in the cultural heritage domain, closely related to digital preservation. Therefore, this keynote speech raised some common issues with PRELIDA objectives, serving the purpose of introducing means of interdisciplinary collaboration to students.

Other talks during the summer school introduced basic aspects of Linked Data and semantic technologies, as well as several challenges that need to be tackled in this domain. For instance, during the Q&A session, questions about permanent URIs and dereference, which have been also discussed during other PRELIDA activities, were posed to the joint panel of tutors.

Regarding the Linked Data Preservation track, the participant tutors presented a handful of challenges and current approaches for the preservation of Linked Data. Students were actively participating during the talks, and especially in the hands-on sessions, where they were able to analyse the existing problems in Linked Data preservation and to propose their own solution ideas.

The first session of this track was presented by Dr. Dumitru Roman [2], and he gave an overview of Data-as-a-Service Solutions, with a focus on Open Data and the DaPaaS Platform². Data-as-a-Service (DaaS) is emerging as a new paradigm for cost-effective and agile data provisioning, promising to simplify data management for organizations with limited expertise in the field, and to reduce the costs for data integration, publishing and consumption. The goal of this presentation was to familiarize the summer school students with the DaaS concept and the emerging solutions in this domain, and to serve as a reference point for those interested in working with DaaS solutions and technologies. In this setting, long-term preservation is a key feature to consider.

Then, Dr. René van Horik introduced research data management as a key activity in scholarly activities [3]. Research data management (RDM) concerns adding value and maintaining access to research data over the course of its scholarly lifecycle. In this session the emerging international infrastructure for RDM was described and discussed, including an overview of tools and services (such as research data centers). Funders, such as the EU Horizon2020 program, increasingly require a RDM plan as part of a project proposal, so this session was focused on showing the relevance and implementation of RDM plans, since datasets are considered as highly valuable assets. In summary, the talk covered five themes: (1) description of the data to be collected / created (i.e. content, type, format, volume...); (2) Applicable standards / methodologies for data collection and management; (3) Ethics and Intellectual Property; (4) Plans for data sharing and (open) access; and (5) strategy for long-term preservation. Students were able to investigate to what extent existing principles and templates for data management plans are applicable to manage Linked Data objects.

The last part of the Linked Data Preservation track was comprised of a tutorial and a hands-on session focused on concrete challenges on preserving Linked Data. First, Dr. Sotirios Batsakis introduced the notion of long-term preservation of linked data, identifying challenges, problems and limitations of current digital preservation approaches when applying them to Linked Data [4]. During his presentation he discussed the current state of the art, focusing on the existing gap between the Linked

² <http://dapaas.eu/>



Data and the Digital Preservation communities, and providing examples of real-world use cases. Then, based on particular use cases, such as DBpedia, Dr. José María García presented a series of practical exercises for the students to analyse and discuss how digital preservation techniques can help to ensure the long-term availability of Linked Data [5]. Finally, during this last session he also introduced which are the technical and organisational issues that need to be taken into account to ensure the required preservation, and analysed which particular measures are put into place in real use cases.

5. On-line Resources

The chapter lists on-line resources created and used for organizing the summer school and describes dissemination activities to promote and engage prospective students.

General ESWC 2014 Summer School website is available at the following address:

<http://summerschool2014.eswc-conferences.org/>

The summer school website includes online learning materials on the summer school's topics from the Educational Curriculum for the Usage of Linked Data (EUCLID) materials' collection channel [6], which were provided to prospective attendees (Figure 1) as pre-reading material in convenient formats as eBook (HTML, iBook, ePub, Kindle) and Course (HTML, iTunesU, Slides, Webinar), available for download at :

<http://summerschool2014.eswc-conferences.org/learning-materials>

Euclid Project Modules

EUCLID Module 1: Introduction and Application Scenarios

This module introduces the main principles of Linked Data, the underlying technologies and background standards. It provides basic knowledge for how data can be published over the Web, how it can be queried, and what are the possible use cases and benefits. As an example, we use the development of a music portal (based on the MusicBrainz dataset), which facilitates access to a wide range of information and multimedia resources relating to music. The module also includes some multiple choice questions in the form of a quiz, screencasts of popular tools and embedded videos.

eBook

HTML iBook ePub Kindle

Course (includes screencasts and exercises)

HTML iTunesU Slides Webinar I Webinar II

Figure 1: Snapshot of Learning Materials section of the ESWC 2014 Public Website.

Noteworthy, PRELIDA co-branded the organization of the event: PRELIDA project logo appeared on the public website, school's flyers and additional dissemination materials provided by the summer school. Moreover, the information about the school has been circulated among the digital preservation communities in order to recruit prospective students.

Additionally, to involve the community and increase dissemination of events related to the summer school, a dedicated Facebook group was created and maintained:

<https://www.facebook.com/groups/eswcsummerschool/>

Presentations from the school were uploaded to Slideshare account of the summer school:

<http://www.slideshare.net/eswcsummerschool/>



6. Conclusion

The major objective of the summer school was to engage young scientists in solving the emerging challenges of Linked Data, Big Data and other Semantic Web related topics. As part of the talks, the PRELIDA project was introduced to students: its major objectives, research questions and current challenges were discussed in keynotes and in specific tutorials and hands-on sessions about Linked Data preservation. Students were not hesitating to ask various related questions, participate in discussions about Digital Preservation and its relations to the Linked Data.

The organization of the school was heavily supported by engagement through the social channels such as the public website, Twitter, Facebook group page, Slideshare etc. All of the talks and presentations were recorded by Videlectures.net³ and will soon be uploaded as videos to the ESWC Summer School 2014 account.

In summary, the second summer school was a major effort to effectively disseminate the preliminary outcomes of the project, presenting the state of the art in Linked Data preservation and the challenges and opportunities present in the area.

³ <http://videlectures.net/>

References

- [1] Dominic Oldman, “Conflicting Cultures of Knowledge”, 2014. Available at <http://www.slideshare.net/eswcsummerschool/wed-oldman-keydifferent-cultures-and-knowledge>
- [2] Dumitru Roman, “Open Data Publication and Consumption”, 2014. Available at <http://www.slideshare.net/eswcsummerschool/wed-roman-tutopendatapub-38742186>
- [3] René van Horik, “Research Data Management”, 2014. Available at <http://www.slideshare.net/eswcsummerschool/wed-van-horikhandsonresearch-data-management-38742188>
- [4] Sotiris Batsakis, “Challenges on Preserving Linked Data”, 2014. Available at <http://www.slideshare.net/eswcsummerschool/wed-batsakis-tutchallenges-of-preservations-38742201>
- [5] José María García, “Preserving Linked Data: Hands-on use cases”, 2014. Available at <http://www.slideshare.net/eswcsummerschool/wed-garcia-handsondbpedia-preservation-38742189>
- [6] EdUcational Curriculum of the usage of Linked Data – EUCLID Project: <http://euclid-project.eu/>