



Project no. 600663

PRELIDA

Preserving Linked Data
ICT-2011.4.3: Digital Preservation

D2.4 Report on the first 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: D2.4
Deliverable title: Report on the first summer school
Due date of deliverable: 08/2013
Actual date of deliverable: 10/2013
Author(s): Serge Tymaniuk, Anna Fensel, Jose Maria Garcia
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
1.0	19/09/2013		Serge Tymaniuk (UIBK)	First draft
1.1	25/09/2013		Serge Tymaniuk (UIBK)	Updated chapters 2-6

Executive Summary

This deliverable reports on the first summer school organized by the PRELIDA project. The school ran in conjunction with the ESWC Summer School 2013 from September, 2nd to September 7th, 2013, in Kalamaki, Greece, and focused on various Linked Data and Digital Preservation topics. Experts from Digital Preservation communities were invited to emphasize the current challenges, outline vital research questions 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 Workshop.....	5
4. Scientific Outcome	8
5. On-line Resources	9
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. Senior experts from Digital Preservation communities both from the Academia as well as from the Research were invited to give keynote talks and provide tutorial and hands-on sessions during the school. One of the senior key experts from the Industry, invited by the summer school, was the development manager of ResearchSpace¹ platform at the British Museum. 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 an expert focusing on Digital Libraries in such fields as personalization and preference management, peer-to-peer data and knowledge management systems, publish/subscribe systems, graph mining and power laws. 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 1st PRELIDA Summer School hold in conjunction with the ESWC 2013 Summer School on September 2-7, 2013 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.

¹ <http://www.researchspace.org/>



2. Participants

The summer school includes 39 students as well as tutors, keynote speakers and an organizing team.

The tutorial team included the following experts:

Maribel Acosta (Karlsruhe Institute of Technology)
Marko Grobelnik (Cycorp Europe, Quintelligence, Jozef Stefan Institute)
John Domingue (The Open University)
Aidan Hogan (Digital Enterprise Research Institute)
Jarred McGinnis (Logomachy Ltd, King's College London)
Barry Norton (The British Museum)
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 field and included the following speakers:

Vassilis Christophides (University of Crete)
Stefan Decker (National University of Ireland)
Kieron O'Hara (University of Southampton)
Steffen Staab (University of Koblenz-Landau)

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

Aneta Tumilowicz
Fabian Flöck

3. Programme of the Workshop

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 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 such as VOID vocabulary, Open Refine, GATECloud, relational databases, NLP and corresponding tools. For instance, SILK Link Discovery framework and SILK link specification language were also discussed.
 - o **Ontologies (2) module** focused on presenting the best practices in ontology engineering: the most successful approaches and frameworks, examples and use cases.
 - o **Big data (3) module** aimed to give an overview of the big data notion, its management, analytics, tools and relations to Linked Data. The most popular programming model, i.e. MapReduce approach was analysed and the most popular implementations were discussed.
- **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, Big Data 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 3-4 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 2

09:00	09:15	John Domingue	Opening session
09:15	10:15	Stefan Decker	Keynote: From Linked Data to Networked Knowledge
		Aidan Hogan, Jarred McGinnis	Tutorial: Introduction to Linked Data: background technologies and standards, motivating application scenario
10:30	12:30		<i>Lunch</i>
12:30	14:00		
		Aidan Hogan, Jarred McGinnis	Tutorial: Querying Linked Data
14:00	15:00		
		Aidan Hogan, Jarred McGinnis	Hands-on: Semantic Web languages and standards
15:00	17:00		Poster session and opening reception, <i>dinner</i>
19:00	-	Elena Simperl	

Tuesday, September 3

9:00	9:15	John Domingue	Administrative issues
9:15	10:15	Steffen Staab	Keynote: Programming the Semantic Web
		Barry Norton, Maribel Acosta	Tutorial: Providing Linked Data
10:30	12:30		<i>Lunch</i>
12:30	14:00		
		Fabian Flöck	Student project kick-off
14:00	14:30		
		Barry Norton, Maribel Acosta	Tutorial continued: Providing Linked Data
14:30	15:30		Hands-on: Publishing and consuming Linked Open Data
15:30	17:00	Maribel Acosta	Poster session continued and <i>dinner</i>
19:00	-	Elena Simperl	

Wednesday, September 4

09:00	09:15	John Domingue	Administrative issues
09:15	10:15	Kieron O'Hara	Keynote: The Information Spring
10:30	11:30	Elena Simperl	Tutorial: Building and using ontologies
11:30	12:30	Elena Simperl	Hands-on: Building and using ontologies
12:30	14:00		<i>Lunch</i>
14:00	15:00	Marko Grobelnik	Tutorial: Big Data management and analytics



15:00 16:00 Marko Grobelnik
16:00 -

Hands-on: Big Data management and analytics
Excursion and dinner

Thursday, September 5

09:00 09:15 John Domingue
 Vassilis
09:15 10:15 Christophides
10:30 11:30 all
11:30 12:30
12:30 14:00
14:00 18:00
20:00 -

Administrative issues
Keynote: Preserving Linked Data
Feedback on teaching materials
Student project work (tutors available)
Lunch
Student project work (tutors available)
Social Dinner

Friday, September 6

09:00 09:15 John Domingue
09:15 10:15 Marko Grobelnik
10:15 12:30
12:30 14:00
14:00 18:00
20:00 -

Administrative issues
Keynote: Big data
Student project work (tutors available upon request)
Lunch
Student project work (tutors available)
Dinner and party

Saturday, September 7

09:00 12:00 John Domingue
12:00 12:30
12:30 13:00

Student project presentations
Break and deliberation of judges
Award ceremony and closing

4. Scientific Outcome

During the keynote of Prof. Vassilis Christophides, major challenges in the Linked Data preservation field were discussed [1]. He introduced digital data notion, preservation as a concept, the motivation of the need of digital objects preservation and their significant properties. In addition, common digital preservations techniques and possible connections to Linked Data were analysed. In his view, one of the prerequisite for preserving the Linked Data is ensuring proper data quality in order to trust the preserved data. Initially, Linked Data preservation heavily relies on data publication and refinement. Thus, it is vital to verify the completeness, singularity, consistency and correctness of online published data. Although, as Vassilis Christophides emphasized, the publishing of data and its preservation were usually thought of as primarily separate processes, however there was a need “to narrow the traditional boundaries between the data creators and publishers and the data archivists and brokers” ([1] p.17). This brings us to several research problems as was stated during the keynote, precisely: *the entity resolution problem*, i.e. finding ways for identifying that resource descriptions across datasets match to the same entity in the real world; *the provenance problem*, i.e. finding ways of recording datasets dependencies; *the evolution problem*, i.e. finding solutions of change monitoring of third-party datasets; and *consistency problem of multi-version archives*, i.e. choosing the most suitable versions of linked datasets for future use preservation.

The speaker proposed the need for the creation of new abstraction required to bring together data creation, publication and preservation. Finally, considering management perspective, as was pointed out during the talk, there existed a need of new business models to be created in order to consider and spread data preservations costs among stakeholders.

Dr. Barry Norton, the development manager at the British Museum, a tutorial and hands-on sessions’ leader and one of the students’ projects coordinator, covered various IT topics relevant to digital management and digital analysis. Based on his experience in cultural heritage domain, during the discussions, he argued for the need to link semantic technologies in the digital environment in order to ensure better integrity, openness and greater availability.

A specific slot of the summer school’s tutorials provided by Dr. Elena Simperl was dedicated to ontology management, which is specifically vital for the cultural heritage domain to ensure proper data preservation. During the tutorial the problems of finding and reusing existing ontologies and Linked Open vocabularies for various domains including the cultural heritage were raised and discussed.

Noteworthy, some of the attending students had already been involved in certain research projects of digital preservation of cultural properties. For example, one of the nominees for the best poster, Gopala Koduri, presented the research poster regarding the development of a system for exploring local folklore audio music exploiting cultural context [2].



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 2013 Summer School website is available at the following address:

<http://summerschool2013.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 [3], 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://summerschool2013.eswc-conferences.org/learning-materials/materials/>

A screenshot of the ESWC 2013 Learning Materials section. At the top, there is a navigation bar with links: Home, About, Tutors, Keynotes, Attending, and Learning Materials (which is highlighted in pink). Below the navigation bar, there is a heading "EUCLID Module 1: [Introduction and Application Scenarios](#)". A text block describes the module's content, mentioning Linked Data, music portals, and various learning formats. Below the text, there are two rows of download links. The first row is labeled "eBook" and includes buttons for HTML, iBook, ePub, and Kindle. The second row is labeled "Course (includes screencasts and exercises)" and includes buttons for HTML, iTunesU, Slides, Webinar I, and Webinar II. A small "II" icon is positioned between the two rows.

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

Course (includes screencasts and exercises)

HTML iBook ePub Kindle

HTML iTunesU Slides Webinar I Webinar II

Figure 1: Snapshot of Learning Materials section of the ESWC 2013 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/eswccsummerschool/>

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

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



During the running of the school the most prominent conclusions were disseminated through various PRELIDA online social channels such as Facebook, Twitter and Google Plus. A total of 19 tweets/posts dedicated to the summer school activities were created.

In addition, the blog post was published to the PRELIDA public blog with the major results from the school.



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. 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 Videolectures.net² and will soon be uploaded as videos to the ESWC Summer School 2013 account.

Overall, the summer school was the first big step in disseminating the PRELIDA research topics to the next generation of scientists and facilitating relevant discussions to aid students in identifying their research interests and aligning it with the modern trends in the Linked Data world.

The second summer school is planned to be held around the PRELIDA Consolidation and Dissemination Workshops and present preliminary PRELIDA roadmaps, collect feedback and discuss opportunities and challenges.

² <http://videolectures.net/>



References

- [1] V.Christophides. Preserving Linked Data: Challenges and Opportunities. Presentation at the ESWC 2013 Summer School, Kalamaki, Greece. University of Crete & FORTH-ICS.
- [2] Ontologies, Semantic Web and CompMusic, Blog post available online: <http://compmusic.upf.edu/node/170>
- [3] EdUcational Curriculum of the usage of Linked Data – EUCLID Project: <http://euclid-project.eu/>