

General
Informations

Date of Birth: 10 September 1986
Place of Birth: Elbasan, Albania
Citizenship: Albania
Address: via Aldini 51, 20157 Milano (MI), Italy
Mobile: +39-324-686-6862
E-mail: spahiu.blerina@gmail.com

Qualifications
and Interests

My research interests involve Semantic Web technologies, Machine Learning, Quality of Linked Data, Data Profiling, Big Data Analytics, Artificial Intelligence.

Current
Employment

DISCO - UNIMIB (Department of Informatics, Univeristy of Milano - Bicocca), Milano, Italy:

- April 2017 - Present
- PostDoc Researcher on the development of tools and approaches for automatic profiling of semantic data.
- Key topics covered: data profiling, knowledge graphs, semantic data, big data, machine learning, knowledge graph construction.

Professional &
External Activities

- Task 1.4 leader in the **COST Action CA18209**: European network for Web-centred linguistic data science.
- External Reviewer for projects of the Esmera Project Call 2020, Reviewing Period: January-February 2020.
- Intensive teaching course for TechRAIN company. A full-day course on NoSQL databases, with a particular focus on MongoDB and querying DBs using Shell and Python libraries.
- Journal Peer Reviewer: ACM Journal of Data and Information Quality 2019, Future Generation Computer Science 2019, Information Systems 2019, etc.
- Member of the international scientific committee of conferences ISWC2017-ISWC2021, ESWC2017-ESWC2021, KEOD 2019, WWW2018, GeoLD2018, LDQ2017, EKAW2016, Semantics2016, KEOD2016, KSEM2015.
- Invited Talk: "Enhancing robots knowledge with Semantic Web Technologies"- In the IEEE Conference on Emerging Technologies and Factory Automation, September 4th - 7th, 2018, Torino, Italy

Previous
Employments

INSTAT (National Institute of Statistics, Tirana, Albania)

- October 2009 - January 2014
- IT Specialist of IT Department
- **Responsible** for Active Directory, Windows Server 2003/2008, Microsoft Exchange 2007- 2010, Data Protection Manager, System Center essential 2010, Network monitoring. Installation of different programs, anti virus management, network management, servers management, VPN management, etc
- Teaching Assistant for Courses: MS Excel and SPSS.

ADHS (Albanian Demographic and Health Survey Project, Tirana, Albania)

- October 2008 - October 2009

- **Responsible** for software and hardware maintenance, programming some of the tabulations for the results of the fieldwork, writing reports, programming some of the tables of the final report for the “Albanian Demographic and Health Survey” (ADHS). Supervising the quality of the data collected from the teams in the field. Working with CsPro Program for data editing, data cleansing and processing.

Research member for the Research Project: The socio-demographic changes and child development. Tirane, Albania

- September 2010 - March 2011
- Running tables for the results of the survey, working in team for analyses of the research and writing the final report.

Scholar Practice Department of IP& DATA and Department of Card Phonic System, Albtelekom Sha, Albania

- March 2008 - June 2008
- Area of Work: IP and Data Management
- Activities: analyze and find solutions to the problems of Card Phonic Systems.

Foreign Languages

- English (TOEFL exam) - Proficiency C2
- Italian (CELI exam) - Proficiency C2
- Albanian - Mother Tongue

Software Skills Java, Python, R, SPARQL, MongoDB, Neo4J, RapidMiner, SPSS, Casandra

Education University Education: **University of Milano - Bicocca**, 20126 Milano (MI), Italy

Ph.D., In Informatics, February 2014 - March 2017

- Thesis Title: Profiling Linked Data
<https://boa.unimib.it/handle/10281/151645>
- Area of Study: informatics
- Ph.D. external period abroad at **University of Mannheim, Germany**.
- The doctoral research tackled problems related to Profiling Linked Data in more details the thesis covered three aspects of profiling; topic-based, schema-based and linkage-based. Such profile is fundamental for the decision-making process and is the basic requirement towards the dataset understanding. An approach to automatically classify datasets in one of the topical categories used in the LD cloud is presented. For the schema-based profiling a schema-based summarization approach is proposed that provides an overview about the relations in the data. The summaries are concise and informative enough to summarize the whole dataset. Moreover, they reveal quality issues and can help users in query formulation tasks. Many datasets in the LD cloud contain similar information for the same entity. In order to fully exploit its potential LD, such information should be made explicit. Linkage profiling provides information about the number of equivalent entities between datasets and reveal possible errors.

Courses and Summer Schools during PhD:

1. Clustering Analyses
2. Ubiquitous and Adaptive Systems: Innovation-Driven Modeling and Design
3. Big Data Profiling
4. ESWC summer school - August 31- September 5, 2015, Crete, Greece
5. ALIAS summer school - September 14-19, 2015, Como, Italy
6. European Data Science Summer School - September 26-30, 2016, Saarbruecken, Germany

MSc, Electronic Engineering, September 2008 - July 2012

- Thesis Title: Performance of different operating systems in VMware
- Area of Study: electronic and informatics
- Key courses: Deepen Studies in Network Administrations, Analogue Electronics, Digital Electronics, Java Programming, Fibre Optics, etc

BSc, Electronic Engineering, September 2005 - September 2008

- Thesis Title: Electronics in medicine: ECG
- Area of Study: electronics and informatics design

High School: Kostandin Kristoforidhi, Elbasan, Albania

- September 2001 - July 2005

Awards

- **Best Paper Award** at WOP co-located with the International Semantic Web Conference 2018, "Towards Improving the Quality of Knowledge Graphs with Data-driven Ontology Patterns and SHACL".
- Selected to participate to **IBM's Best Student Recognition Event 2016**
- **Best Paper Award** at SUMPRES co-located with the European Semantic Web Conference 2016, "ABSTAT: Ontology-driven Linked Data Summaries with Pattern Minimalization".

Teaching Experience

Teaching Assistant

- Assistant for Database Systems, Department of Informatics, University of Milan - Bicocca, Italy. Organize laboratories and exercises for students for the course of Database Systems of the Bachelor in Informatics (2019-2021).
- Assistant for the Data Semantics, Department of Informatics, University of Milan - Bicocca, Italy. Organize exercises and support students for the course of Data Semantics of the Master in Data Science (2018-2020).
- Assistant for the Data Technology, Department of Informatics, University of Milan - Bicocca, Italy. Organize exercises and support students for the course of Data Technology of the Master in Informatics (2018-2020).
- Laboratory assistant: Programming in Java Tutor, Department of Informatics, University of Milan - Bicocca, Italy. Support first years students of Bachelor in Informatics in the laboratories for Java Programming Language (2016-2020).
- External Lecturing Professor. Lecturing and seminars of the subject "Theory of Research" at the Faculty of Social Science, University of Tirana, Albania for the Bachelor in Informatics (2012-2014).
- External Lecturing Professor. Lecturing and seminars of the subject "Information System Development" and "Computer Applications" at the Faculty of Information Technology, University of Durres, Albania (2012-2014).

Projects

Nexus Linguarum eCOST Action

The main aim of this Action will be to promote synergies across Europe between linguists, computer scientists, terminologists, language professionals, and other stakeholders in industry and society, in order to investigate and extend the area of linguistic data science.

- task 1.4 leader
- profile and evaluate the quality of linguistics data.
- development of quality metrics for linguistic data.

Food Social Sensor Network (FODNET), University of Milan - Bicocca, Italy.

Food Net is a Smart Cities and Communities project that aims to: identify and interpret the nutritional needs of specific consumer targets (citizens over 65 of the metropolitan city), encourage process and product innovation of agri-food companies through implementation

of infrastructures that allow to create truly effective functional foods, implement the quality of life and the nutritional status of citizens through functional food, nutritional models and educational processes aimed at improving lifestyle. Duration 2018-2021.

- main goals: propose and develop an ontology for food and nutrition needs.
- profile and evaluate the quality of food data.

European Business Graph (euBusinessGraph), University of Milan - Bicocca, Italy.

The project aims to create a cross-border knowledge graph of companies, and a set of innovative business products and services build upon the knowledge graph. Duration 2018-2019.

- main contribution in defining and developing an ontology for event data.

Supporting Event and Weather-based Data Analytics and Marketing along the Shopper Journey (EW-SHOPP), University of Milan - Bicocca, Italy.

The aim of the project is to build a platform with innovative services based on advanced data analysis techniques capable to leverage the large variety and volume of data from multiple sources. Duration 2018-2019.

- definition and development of a reconciliation services for data enrichment purposes.

Components Supporting The Open Data Exploitation Project (COMSODE Project), University of Milan - Bicocca, Italy.

The aim of the project is to develop a methodology to support the public administration in the publication of Open Data. Duration 2014-2016.

- the main focus is writing a methodology on how to make data profiling and quality assessment/improvement of the datasets.

Journal
Publications

- [1] Renzo Arturo Alva Principe, Andrea Maurino, Matteo Palmonari, Michele Ciavotta, **Blerina Spahiu**: ABSTAT-HD: A scalable tool for profiling very large Knowledge Graphs". In: The International Journal on Very Large Data Bases - To appear
- [2] Maria Pia di Buono, Hugo Gonçalo Oliveira, Verginica Barbu Mititelu, **Blerina Spahiu**, Gennaro Nolano. Paving the Way for Enriched Metadata of Linguistic Linked Data. Semantic Web Journal, under second review (minor revision).
- [3] Marco Cremaschi, Flavio De Paoli, Anisa Rula, **Blerina Spahiu**: A fully automated approach to a complete Semantic Table Interpretation. Future Generation Computer Systems 112: 478-500 (2020).
- [4] **Blerina Spahiu**, Andrea Maurino and Robert Meusel: Topic profiling benchmarks in the Linked Open Data Cloud: Issues and lessons learned". In: Semantic Web Journal 10.2, pp. 329-348 (2019).
- [5] Gianluigi Viscusi, **Blerina Spahiu**, Andrea Maurino, Carlo Batini. Compliance with open government data policies: An empirical assessment of Italian local public administrations. In Information Polity Journal 19.3-4, pp. 263-275.

Conferences and
Workshop
Publications

- [6] Marco Valzelli, Andrea Maurino, Matteo Palmonari and **Blerina Spahiu**: Towards an Access Control Model for Knowledge Graphs. SEBD 2021. To appear.
- [7] Michele Ciavotta, Vincenzo Cutrona, Flavio De Paoli, Matteo Palmonari, **Blerina Spahiu**: "BEEO: Semantic Support for Event-based Data Analytics". In the International Semantic Web Conference (ISWC) 2021. To appear.
- [8] Giuseppe De Pasquale, **Blerina Spahiu**, Pietro Ducange, Andrea Maurino: Towards Automatic Classification of Sheet Music. SEBD 2020: 266-277.

- [9] Loris Roveda, **Blerina Spahiu**, Walter Terkaj. “On the Proposal of a Unified Safety Framework for Industry 4.0 Multi-Robot Scenario”. In: Proceedings of the 27th Italian Symposium on Advanced Database Systems, Castiglione della Pescaia (Grosseto), Italy, June 16-19, 2019
- [10] **Blerina Spahiu**, Andrea Maurino, Matteo Palmonari. “Towards Improving the Quality of Knowledge Graphs with Data-driven Ontology Patterns and SHACL”. **BEST PAPER AWARD!** Proceedings of the 9th Workshop on Ontology Design and Patterns (WOP 2018) co-located with 17th International Semantic Web Conference (ISWC 2018), Monterey, USA, October 9th, 2018. Pp. 52–66
- [11] **Blerina Spahiu**, Vincenzo Cutrona, Matteo Palmonari. “Towards a Unified Geospatial Data Reconciliation Platform for Data Analytics”. In: Proceedings of the 2nd Mobility Analytics for Spatio-temporal and Social Data co-located with 44th International Conference on Very Large Data Bases VLDB 2018 Rio De Janeiro, Brazil, 27th-31st August, 2018.
- [12] Renzo Arturo Alva Principe, **Blerina Spahiu**, Matteo Palmonari, Anisa Rula, Flavio DePaoli, Andrea Maurino. “ABSTAT 1.0: Compute, Manage and Share Semantic Profiles of RDF Knowledge Graphs”. In: Proceedings of the 15th Extended Semantic Web Conference ESWC 2018, Crete, Greece, 3rd-7th June, 2018
- [13] Renzo Arturo Alva Principe, **Blerina Spahiu**, Matteo Palmonari, Anisa Rula, Flavio De Paoli, Andrea Maurino: ABSTAT 1.0: Compute, Manage and Share Semantic Profiles of RDF Knowledge Graphs. ESWC (Satellite Events) 2018: 170-175
- [14] **Blerina Spahiu**, Cheng Xie, Anisa Rula, Andrea Maurino, Hongming Cai: Profiling Similarity Links in Linked Open Data. In: Proceedings of the 7th International Workshop on Data Engineering meets the Semantic Web In conjunction with ICDE 2016, Helsinki, Finland, May 16th, 2016
- [15] **Blerina Spahiu**, Riccardo Porrini, Matteo Palmonari, Anisa Rula, Andrea Maurino: AB-STAT: Ontology-driven Linked Data Summaries with Pattern Minimalization”. **BEST PAPER AWARD!** Proceedings of the 2nd International Workshop on Summarizing and Presenting Entities and Ontologies. In conjunction with ESWC2016, in Anissaras, Crete, Greece, May 29th, 2016 to June 2nd 2016.
- [16] Cheng Xie, Melisachew Chekol, **Blerina Spahiu**, Christian Meilicke, Hongming Cai. “Leveraging Structural Information in Ontology Matching”. In: The 30-th IEEE International Conference on Advanced Information Networking and Applications (AINA-2016), Switzerland, March 23-25, 2016.
- [17] Cheng Xie, Dominique Ritze, **Blerina Spahiu**, Hongming Cai: Instance-based property matching in linked open data environment”. In: Proceedings of the Ontology Matching in the 14th ISWC 2015, Pennsylvania, US, Oct 11-15, 2015.
- [18] Matteo Palmonari, Anisa Rula, Riccardo Porrini, Andrea Maurino, **Blerina Spahiu**, Vincenzo Ferme: ASBTAT: Linked Data Summaries with ABstraction and STATistics. In: The Semantic Web: ESWC 2015 Satellite Events - ESWC 2015 Portoroz, Slovenia, May 31th, 2015 to June, 4th, 2015.
- [19] Robert Meusel, **Blerina Spahiu**, Christen Bizer, Heiko Paulheim. “Towards Automatic Classification of LOD datasets”. In: Proceedings of the Workshop on Linked Data on the Web co-located with the 24th International World Wide Web Conference (WWW 2014), Firenze, Italy, May 19, 2015.
- [20] **Blerina Spahiu**. “Profiling Linked Open Data”. In: Proceedings of the Doctoral Consortium in the 14th International Semantic Web Conference (ISWC 2015), Pennsylvania,

US, Oct 11-15, 2015 and 14th Conference of the Italian Association for Artificial Intelligence 2015 Italy.

- [21] Gianluigi Viscusi, **Blerina Spahiu**, Andrea Maurino, Carlo Batini. Compliance with open government data policies: An empirical assessment of Italian local public administrations. In: Rethinking Information Systems in the Public Sector, 8 June 2014, Tel Aviv.

Personal Data

I authorize the use of my personal data in accordance with Legislative Decree 196 of June 30 2003

Consapevole delle responsabilità penali in cui può incorrere in caso di dichiarazione mendace, dichiaro che il contenuto del curriculum costituisce dichiarazione ai sensi della normativa vigente in tema di autocertificazioni