

The 1st edition of the Data Quality meets Machine Learning and Knowledge Graphs (DQMLKG) workshop was held in Hersonissos, Crete, on May 26th, 2024, co-located with the 21st Extended Semantic Web Conference (ESWC 2024).

The DQMLKG workshop aimed to explore the intricate interplay of data quality, Machine Learning (ML), and Knowledge Graphs (KGs). This includes elucidating limitations in assessment methodologies, proposing effective methods for objective quality assessment, and addressing challenges on ML and Artificial Intelligence (AI) in general, as well as as verifying if and to what extent well-known quality metrics are compliant with ML-based quality assessment, and addressing FAIR principles. We also welcomed proposals riding the path of Explainable AI, Large Language Models, Generative AI, and any AI-driven approach that can be applied to Semantic Web technologies to support and enhance data quality assessment and improvement.

The topics of interest included the following:

- New approaches for performing Data quality assessment or improvement of Knowledge Graphs via Machine Learning
 - Quality assessment over time
 - Scalability issues
 - Proactive approaches able to improve KG quality during the data authoring stage
 - Reactive approaches to improve KG quality before the data exploitation stage
 - Large Language Models to deal with KG quality issues
 - Generative Artificial Intelligence (AI) to cope with KG quality issues
 - AI-driven approach to assess and improve data quality issues over KGs
- Applications combining Machine Learning and Knowledge Graphs dealing with Data Quality concerns:
 - Recommender Systems leveraging (incomplete) Knowledge Graphs
 - Link Prediction and completing KGs
 - Ontology Learning and Matching coping with KG consistency and accuracy
 - Question Answering exploiting Knowledge Graphs and Machine Learning dealing with representational issues
 - Domain Specific KG quality issues

DQMLKG 2024 opened with an exciting keynote entitled “When stars align: studies in data quality, knowledge graphs, and machine learning” held by Elena Simperl, Professor of Computer at King’s College London and Director of Research for the Open Data Institute. She emphasized that Data quality is *the* topic all of us should care about.

Then, the workshop led to the presentation of four contributions from researchers across Europe and mainly concerned the use of Large Language Models for supporting quality assessment, the impact of Linguistic KGs in dealing with formality, and how to extend shape expressions to approach different types of KGs. All contributions were presented in person at the workshop in Hersonisso, Crete. This encouraged an interesting and engaging discussion and a profitable questions-and-answers session. Finally, the discussion continued with a panel composed of Elena Simperl, Paul Groth, Heiko Paulheim, and Anastasia Dimou with Mehwish Alam as a moderator. This provided all participants with insights about the future of the topic and strongly suggested investing in this direction.

All contributions have been published on <http://ceur-ws.org> as joint proceedings with the 3rd International Workshop on Knowledge Graph Generation from Text (TEXT2KG). At least two program committee members reviewed the submissions of DQMLKG.

We want to thank the program committee members, the authors, and all the participants.

The organizers,

Maria Angela Pellegrino (University of Salerno, ITALY),
Anisa Rula (University of Brescia, ITALY),
Jose Emilio Labra Gayo (University of Oviedo, SPAIN),
Michael Cochez (Vrije Universiteit Amsterdam, THE NETHERLANDS),
and Mehwish Alam (Institut Polytechnique de Paris, FRANCE).

Program committee:

- Cinzia Cappiello, Polytechnic of Milan, Italy
- Jeremy Debattista, Trinity College Dublin, Ireland
- Anastasia Dimou, Katholieke Universiteit, Leuven
- Paul Groth, University of Amsterdam, Holland
- Antonio Lieto, University of Salerno
- Ernesto Jiménez-Ruiz, University of London, England
- Blerina Spahiu, University of Milan, Italy