





# Preface of MEPDaW 2021: Managing the Evolution and Preservation of the Data Web

Fabrizio Orlandi<sup>1</sup> , Damien Graux<sup>2</sup> ,  
Julio Cesar dos Reis<sup>3</sup> , and Maria-Esther Vidal<sup>4</sup> 

<sup>1</sup> ADAPT SFI Centre, Trinity College Dublin, Ireland. [orlandif@tcd.ie](mailto:orlandif@tcd.ie)

<sup>2</sup> Inria, Université Côte d'Azur, CNRS, I3S, France. [damien.graux@inria.fr](mailto:damien.graux@inria.fr)

<sup>3</sup> Inst. of Computing, Univ. of Campinas (UNICAMP), Brazil. [jreis@ic.unicamp.br](mailto:jreis@ic.unicamp.br)

<sup>4</sup> Technische Informationsbibliothek (TIB), Germany. [mvidal@umiacs.umd.edu](mailto:mvidal@umiacs.umd.edu)

**Abstract.** The MEPDaW workshop series targets one of the emerging and fundamental problems of the Web, specifically the management and preservation of evolving knowledge graphs. During the past seven years, the workshop series has been gathering a community of researchers and practitioners around these challenges. To date, the series has successfully published more than 35 articles allowing more than 50 individual authors to present and share their ideas.

This 7<sup>th</sup> edition, virtually co-located with the International Semantic Web Conference (ISWC 2021), gathered the community around six research publications and one invited keynote presentation. The event took place online on the 25<sup>th</sup> of October, 2021.

**Keywords:** Web Data evolution · Data preservation, provenance and lineage · Temporal & Evolving Knowledge Graphs · RDF archiving and versioning

## Managing the Evolution and Preservation of the Data Web

There is a vast and rapidly increasing quantity of scientific, corporate, government, and crowd-sourced data openly published on the Web. Open Data plays a catalyst role in the way structured information is exploited on a large scale. A traditional view of digitally preserving these datasets by “pickling and locking them away” for future use, like groceries, conflicts with their evolution. There are several approaches and frameworks (*e.g.* Linked Data Stack [7], Pool-Party Suite<sup>1</sup>, Metaphactory<sup>2</sup>, etc.) targeted at managing the life-cycle of the Data Web. More specifically, these solutions are expected to tackle major issues such as the synchronisation problem (monitoring changes) [9,14], the curation problem (repairing data imperfections) [11], the appraisal problem (assessing the quality of a dataset) [8], the citation problem (how to cite a particular version of a dataset) [12], the archiving problem (retrieving a specific version of a

<sup>1</sup> <https://semantic-web.com/poolparty-semantic-suite/>

<sup>2</sup> <https://metaphacts.com/>

dataset) [10,13], and the sustainability problem (preserving at scale, ensuring long-term access) [12].

The **seventh** edition of this workshop was organised for the second time at the International Semantic Web Conference (ISWC) and followed the structure of the previous editions. We invited a number of experts in the field of Linked Data and Data Evolution & Preservation in order to suggest and advise on the different topics that our workshop covered this year. This year, at ISWC 2022, we successfully gathered more than 50 participants for our half-day event. In line with most academic events, this year MEPDaW was held as a virtual event and we had to re-think the interactions between participants.

## MEPDaW Scientific programme

The workshop started with the keynote entitled “How can we fix the Web of Data?” given by Prof. Katja Hose<sup>3</sup> from the Department of Computer Science of the Aalborg University (Denmark). She initiated her presentation from the observation that Semantic Web practitioners typically consider the Web of Data as a static corpus of information always available and unmutable; however, “in real life settings”, a broad range of problems hits the practitioners such as unavailability of entire knowledge graphs or dead-links for the associated SPARQL endpoints. And more generally, the current Semantic Web tools and paradigms (almost-) completely miss the concept of versioning and provenance of metadata. During her keynote, Professor Hose highlighted some of the solutions her group developed to mitigate these problems. She first showed how to keep knowledge available for continuous and scalable querying. Then, she presented the attendees an approach that enables community-driven updates so that mistakes can be corrected or missing information can be added. And finally, she described how learning from RDF archiving can be done using solutions to better support evolving knowledge graphs. Overall, this keynote [2] gave the audience in-depth details on practical (and industrial) use cases backed by cutting-edge research techniques.

The first article presented dealt with an approach which helps SPARQL practitioners to know which SPARQL endpoints has been updated when they run complex pipelines relying on several RDF sources [1]. It was followed by [5] which proposes the use of a visual interface to explore and fix multi-dimensional metadata bases, in particular she showed how she will apply these ideas in the context of popular music data during her PhD. Finally, the first paper-session ended-up with the presentation of TrieDF [3]: a solution to index metadata-augmented RDF datasets inspired by the trie data structure.

The second session started with an industrial talk from J. Fernández who described how clinical data standards at Roche benefit from RDF version management. The next effort [6] focused on provided the audience with several application use-cases where our the efforts of our community could contribute to.

---

<sup>3</sup> [http://people.cs.aau.dk/~khose/About\\_me.html](http://people.cs.aau.dk/~khose/About_me.html)

Finally, the last article of the workshop described UpLOD [4], a tool to repair broken links in the linked-open data.

### **Organizing Committee**

- Fabrizio Orlandi, ADAPT SFI Centre, Trinity College Dublin, Ireland
- Damien Graux, Inria, Université Côte d’Azur, CNRS, I3S, France
- Julio Cesar dos Reis, Inst. of Computing, Univ. of Campinas, Brazil
- Maria-Esther Vidal, TIB, Hannover, Germany

### **Advisory Board**

- Philippe Cudré-Mauroux, eXascale Infolab, Univ. of Fribourg, Switzerland
- Jeremy Debattista, TopQuadrant Inc
- Javier D. Fernández, Information Architect at Roche, Switzerland
- Fabien Gandon, Inria, Université Côte d’Azur, CNRS, I3S, France
- Axel Polleres, Vienna University of Economics and Business, Austria

### **Programme Committee**

- Natanael Arndt, Leipzig University, Germany
- Ioannis Chrysakis, FORTH-ICS, Greece; and Ghent Univ. - imec, Belgium
- Pieter Colpaert, Ghent University, Belgium
- Marcos Da Silveira, LIST, Luxembourg
- Christophe Debruyne, Trinity College Dublin, Ireland
- Javier D. Fernández, F. Hoffmann-La Roche AG, Switzerland
- Luis Ibanez-Gonzalez, University of Southampton, England
- Pavel Klinov, Stardog Union, Germany
- Pierre Maillot, Inria, France
- Harshvardhan J. Pandit, ADAPT Centre - Trinity College Dublin, Ireland
- George Papastefanatos, IMIS / RC “Athena”, Greece
- Iliana Petrova, Inria, France
- Fatiha Saïs, LRI & Paris Saclay University, France
- Ruben Taelman, Ghent University – imec, Belgium

### **Acknowledgements**

We would like to thank all the authors, reviewers, committee members and the speakers for their contributions, support and commitment.

These research activities were conducted with the financial support of the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Grant Agreement No. 713567 at the ADAPT SFI Research Centre at Trinity College Dublin. The ADAPT SFI Centre for Digital Media Technology is funded by Science Foundation Ireland through the SFI Research Centres Programme and is co-funded under the European Regional Development Fund (ERDF) through Grant #13/RC/2106\_P2.

## Articles presented at MEPDaW 2021

1. Graux, D., Orlandi, F., O’Sullivan, D.: De-icing federated SPARQL pipelines: a method for assessing the “freshness” of result sets. In: Proceedings of the 7th Workshop on Managing the Evolution and Preservation of the Data Web (MEPDaW) (2021)
2. Hose, K.: Knowledge Graph (R)Evolution and the Web of Data. In: Proceedings of the 7th Workshop on Managing the Evolution and Preservation of the Data Web (MEPDaW) (2021)
3. Pelgrin, O., Galárraga, L., Hose, K.: TrieDF: Efficient in-memory indexing for metadata-augmented RDF. In: Proceedings of the 7th Workshop on Managing the Evolution and Preservation of the Data Web (MEPDaW) (2021)
4. Regino, A., de Jesus Pontes Monteiro, E., dos Santos, A.C., Reis, J.C.D.: UpLOD: A tool for inconsistent links repairment in the LOD. In: Proceedings of the 7th Workshop on Managing the Evolution and Preservation of the Data Web (MEPDaW) (2021)
5. Tikat, M., Winckler, M., Buffa, M.: Interactive multimedia visualization for exploring and fixing a multi-dimensional metadata base of popular musics. In: Proceedings of the 7th Workshop on Managing the Evolution and Preservation of the Data Web (MEPDaW) (2021)
6. Waterman, K.K.: Don’t stop thinking about tomorrow: Use cases demonstrating the asymmetric impact of contextual temporal links in knowledge graph evolution. In: Proceedings of the 7th Workshop on Managing the Evolution and Preservation of the Data Web (MEPDaW) (2021)

## References

7. Auer, S., Böhmann, L., Dirschl, C., Erling, O., Hausenblas, M., Isele, R., Lehmann, J., Martin, M., Mendes, P.N., Van Nuffelen, B., et al.: Managing the life-cycle of linked data with the LOD2 stack. In: International semantic Web conference. pp. 1–16. Springer (2012)
8. Debattista, J., Auer, S., Lange, C.: Luzzu—a methodology and framework for linked data quality assessment. *J. Data and Information Quality* **8**(1) (Oct 2016)
9. Endris, K.M., Faisal, S., Orlandi, F., Auer, S., Scerri, S.: Interest-based RDF update propagation. In: Proceedings of the 14th International Conference on The Semantic Web - ISWC 2015 - Volume 9366. p. 513–529. Springer-Verlag, Berlin, Heidelberg (2015)
10. Fernández, J.D., Polleres, A., Umbrich, J.: Towards efficient archiving of dynamic linked open data. In: MEPDaW workshop at ESWC’15 (2015)
11. Freitas, A., Curry, E.: Big data curation. In: *New Horizons for a Data-Driven Economy* (2016)
12. Gleim, L., Decker, S.: Timestamped URLs as persistent identifiers. In: Proceedings of the 6th Workshop on Managing the Evolution and Preservation of the Data Web (MEPDaW) (2020)
13. Pelgrin, O., Galárraga, L., Hose, K.: Towards fully-fledged archiving for RDF datasets. *Semantic Web (Preprint)*, 1–24 (2020)
14. Tasnim, M., Collarana, D., Graux, D., Orlandi, F., Vidal, M.E.: Summarizing entity temporal evolution in knowledge graphs. In: Companion Proceedings of The 2019 World Wide Web Conference. p. 961–965. WWW ’19, Association for Computing Machinery, New York, NY, USA (2019)