

# Preface of MEPDaW 2018: Managing the Evolution and Preservation of the Data Web\*

Jeremy Debattista<sup>1</sup>, Javier D. Fernández<sup>2</sup>, Jürgen Umbrich<sup>2</sup>, and Maria-Esther Vidal<sup>3</sup>

<sup>1</sup> Adapt Centre, Trinity College Dublin, Ireland  
debattij@tcd.ie

<sup>2</sup> Vienna University of Economics and Business, Vienna, Austria  
{javier.fernandez, juergen.umbrich}@wu.ac.at

<sup>3</sup> Technische Informationsbibliothek (TIB)  
maria.vidal@tib.eu

The 4th edition of this workshop has targeted one of the emerging and fundamental problems in the Semantic Web, specifically the preservation of evolving linked datasets. This topic is of particular relevance to the Semantic Web community since it raises awareness of the many research challenges for preserving and managing dynamic linked datasets. Fostering active usage of such evolving datasets requires further research advances on topics such as storage, synchronisation, change representation and querying over evolving graphs.

This year, we accepted three papers, we invited a keynote speaker, and we discussed on future steps of the community, which we describe in brief.

In this year's contributions we see a focus on the management of data versioning and the preservation of evolving knowledge. Singh et al. [4] present DELTA-LD, a change detection mechanisms for linked datasets. DELTA-LD focuses on detecting changes at both resource level (creation, removal, update, movement, or renewal of a resource) and triple level (deleting or adding a triple). To do so, the approach considers (i) the extraction of features from the linked datasets in order to detect changes and identify similar representations in different versions (i.e. moved resources), and (ii) a classification of the changes and a representation of the change model using a provided ontology.

Pandit et al. [3] investigate on how to represent changes in consents and activities regarding the novel General Data Protection Regulation (GDPR). In their position paper, they first discuss the use of PROV to represent the provenance of activities and ODRL to represent the consent, and identify the influence of consent changes. Then, they discuss on detecting and representing change in activities and how to link and use the changes to demonstrate the compliance w.r.t DDPR obligations.

Laajimi et al. [2] focus on evaluating the performance of archiving engines. In particular, they propose and evaluate the use of the SPARK distributed system to archive RDF data. Thus, authors represent RDF data and changes in SPARK dataframes, while archiving queries are resolved via SPARK SQL. Then, the performance of different versioning approaches (e.g. fully materialized version versus representing only the changing triples in each version) are evaluated, with particular attention to measuring the different performance of star- and chain queries.

---

\* Joint proceedings are publicly available in [1].

Furthermore, in this workshop, Miel Vander Sande keynote<sup>4</sup> provides an in-depth review on the current state of affairs, lessons and challenges when preserving Linked Data. His talk shows, first of all, the importance of preservation (e.g. for link preservation and content and concept drift management) and its roots in Web archiving. In this regards, current strategies can be seen from an observational point of view (discrete snapshots) or being perceived as a continuous flow where the changes can be provided/detected by different ways (e.g. via versioning system or notification-based). In the talk, Miel Vander Sande shows that, although there are (increasingly popular) technical solutions (i.e. products), the Linked Data archiving process (with some building blocks such as interfaces, change detection, publishing, crawling and querying) is still a challenge from the technical but also infrastructural and societal point of view. Finally, in spite of current challenges (e.g. who will be responsible for archiving, parallel truths, robust links), the talk looks at the bright side showing that Linked Data is, in essence, feasible to archive (e.g. it is raw, self-contained and machine processable) and there are many technologies than can support archiving (e.g. Memento, HDT, Triple Pattern Fragments), although we need archiving to be add to the discussion, and stimulate archiving.

We close the workshops with a discussion on open topics and future directions. The main topic was the lack of clear guidelines to archive/preserve linked data, which is preventing further adoption by publishers and consumers. To solve this, we promote the collection of best practices and tools. In addition, we highlight that current systems do not consider the versioned datasets in their query planning strategies. Including this topic in the discussion would foster and boost the performance of current RDF archiving engines.

## Acknowledgments

We would like to thank the authors for their contribution and active participation in the workshops, and all the program committee members for reviewing the submissions and provide valuable feedback. We are also grateful to the organisers of the ESWC 2018 conference for their support, and our keynote speaker, Miel Vander Sande from imec-Ghent University.

The MEPDaW workshop was co-organised by members funded by the EU's Horizon 2020 research and innovation programme: grant 731601 (SPECIAL) and the Austrian Research Promotion Agency's (FFG) program "ICT of the Future": grant 861213 (CitySpin).

## References

1. O. Beyan, J. Debattista, S. Decker, J. D. Fernández, A. Hasnain, M. I. Ali, P. Patel, D. Rebbholz-Schuhmann, D. T. Amit Sheth, J. Umbrich, and M.-E. Vidal, editors. *Joint proceedings*

---

<sup>4</sup> See slides at: <https://mielvds.github.io/MEPDaW2018/>

- of the 4th Workshop on Managing the Evolution and Preservation of the Data Web (MEP-DaW), the 2nd Workshop on Semantic Web solutions for large-scale biomedical data analytics (SeWeBMeDA), and the Workshop on Semantic Web of Things for Industry 4.0 (SWeTI), number 2112 in CEUR Workshop Proceedings, Aachen, 2018.*
2. A. B. Meriem Laajimi and N. Y. Ayadi. Evaluation of rdf archiving strategies with spark. In *4th Workshop on Managing the Evolution and Preservation of the Data Web (MEPDaW)*, 2018.
  3. H. J. Pandit, D. O'sullivan, and D. Lewis. Gdpr-driven change detection in consent and activity metadata. In *4th Workshop on Managing the Evolution and Preservation of the Data Web (MEPDaW)*, 2018.
  4. A. Singh, R. Brennan, and D. O'sullivan. Delta-ld: A change detection approach for linked datasets. In *4th Workshop on Managing the Evolution and Preservation of the Data Web (MEPDaW)*, 2018.