Preface of MEPDaW 2022: Managing the Evolution and Preservation of the Data Web

Damien Graux[1], Fabrizio Orlandi[2], Emetis Niazmand[3,4], Gabriela Ydler[3,4] and Maria-Esther Vidal[3,4]

[1]*Inria, Université Côte d'Azur, CNRS, I3S, France*
[2]*ADAPT SFI Centre, Trinity College Dublin, Ireland*
[3]*Leibniz Information Centre for Science and Technology University Library (TIB), Germany*
[4]*Leibniz University of Hannover, Germany*

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 eight 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 40 articles allowing more than 50 individual authors to present and share their ideas.
This 8th edition, virtually co-located with the International Semantic Web Conference (ISWC 2022), gathered the community around six research publications. The event took place online on the 23rd of October, 2022.

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, governmental, 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], PoolParty Suite[1], Metaphactory[2], 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) [8, 9], the curation problem (repairing data imperfections) [10], the appraisal problem (assessing the quality of a dataset) [11], the citation problem (how to

Managing the Evolution and Preservation of the Data Web (MEPDaW 2022)
damien.graux@inria.fr (D. Graux); orlandif@tcd.ie (F. Orlandi); Emetis.Niazmand@tib.eu (E. Niazmand)
https://dgraux.github.io/ (D. Graux); https://fabriziorlandi.net/ (F. Orlandi)
0000-0003-3392-3162 (D. Graux); 0000-0001-9561-4635 (F. Orlandi); 0000-0001-8194-8079 (E. Niazmand); 0000-0003-1160-8727 (M. Vidal)
© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).
CEUR Workshop Proceedings (CEUR-WS.org)

[1]https://semantic-web.com/poolparty-semantic-suite/
[2]https://metaphacts.com/

cite a particular version of a dataset) [12], the archiving problem (retrieving a specific version of a dataset) [13, 14], and the sustainability problem (preserving at scale, ensuring long-term access) [12].

The **eighth** edition of this workshop was organised for the third 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 "Querying the Web of data using sometimes available APIs" given by Prof. Pieter Colpaert[3] from the IDLab at Ghent University in Belgium. In a context where keeping alive public APIs is complex and where short-time funded projects often turn off their APIs after the end of the project, it is usually complicated to maintain Web applications. In his presentation, Prof. Colpaert described the Linked Data Event Streams which advocates a well preservable API structure (using hypermedia and announcement *via* a metadata catalogue) and provides a way to easily re-deploy data through an API of choice. Overall, this keynote [1] gave the audience in-depth details on practical use cases backed by cutting-edge research techniques.

The first article presented dealt with modelling and analyzing changes within Linked-Data source data [4]. It was followed by by [5] which proposed a method do event sourcing within the SOLID ecosystem.

The second session started with a second keynote from Prof. Aidan Hogan[4], from the University of Chile (Santiago, Chile). His presentation named "Fostering a Lively and Tenacious Web of Data" [2], highlighted key challenges relating to dynamics on the Web of Data: inertia (being slow to change) and impermanence (losing track of the past). He exemplified issues for the Web of Data that may arise if such challenges are left neglected: stale or forgotten data, incorrect results, unchecked vandalism, biased conclusions…Prof. Hogan then discussed research lines to address such challenges relating to representations, modelling, prediction, revision, synchronisation and preservation; and he identified key trade-offs to transition towards a more lively and tenacious Web of Data.

This was followed by the presentations of the last two articles for this year's edition. Nasim *et al.* [3] focused on examining the concept of identity and the notion of redirection within the LOD cloud. Finally, Vercruysse and colleagues [6] described a network of live datasets with the SDS vocabulary.

[3] https://pietercolpaert.be/#me
[4] https://aidanhogan.com/

Organization

Organizing Committee

- Damien Graux, Inria, Université Côte d'Azur, CNRS, I3S, France
- Fabrizio Orlandi, ADAPT SFI Centre, Trinity College Dublin, Ireland
- Emetis Niazmand, Leibniz Universität Hannover & TIB, Germany
- Gabriela Ydler, L3S Forshungszentrum, Research Center, Germany
- Maria-Esther Vidal, Leibniz Universität Hannover & TIB, 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

Programme Committee

- Natanael Arndt, Eccenca GmbH, Germany
- David Chaves-Fraga, Universidad Politécnica de Madrid, Spain
- Ioannis Chrysakis, FORTH-ICS, Greece; and Ghent Univ. – imec, Belgium
- Pieter Colpaert, Ghent University, Belgium
- Marcos Da Silveira, LIST, Luxembourg
- Christophe Debruyne, Université de Liège, Belgium
- Javier D. Fernández, F. Hoffmann-La Roche AG, Switzerland
- Pavel Klinov, Stardog Union, Germany
- Harshvardhan J. Pandit, ADAPT Centre - Trinity College Dublin, Ireland
- George Papastefanatos, IMIS / RC "Athena", Greece
- Iliana Petrova, Inria, France
- Philipp D. Rohde, TIB Leibniz Information Centre, Germany
- 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.

Fundings
These activities were conducted with the financial support of 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. Emetis Niazmand was funded by the EraMed project P4-LUCAT (GA No. 53000015) and the EU H2020 RIA funded project CLARIFY with grant agreement No 875160. Lastly, Maria-Esther Vidal is partially supported by the Leibniz Association in the program "Leibniz Best Minds: Programme for Women Professors", project TrustKG-Transforming Data in Trustable Insights with grant P99/2020.

Articles presented at MEPDaW 2022

[1] Colpaert, P.: Building materializable querying interfaces with the TREE hypermedia specification. In: Proceedings of the 8th Workshop on Managing the Evolution and Preservation of the Data Web (MEPDaW) (2022)
[2] Hogan, A.: Fostering a Lively and Tenacious Web of Data. In: Proceedings of the 8th Workshop on Managing the Evolution and Preservation of the Data Web (MEPDaW) (2022)
[3] Nasim, I., Wang, S., Raad, J., Bloem, P., van Harmelen, F.: What does it mean when your URIs are redirected? Examining identity and redirection in the LOD cloud. In: Proceedings of the 8th Workshop on Managing the Evolution and Preservation of the Data Web (MEPDaW) (2022)
[4] Randles, A., O'Sullivan, D.: Modelling & Analyzing Changes within LD Source Data. In: Proceedings of the 8th Workshop on Managing the Evolution and Preservation of the Data Web (MEPDaW) (2022)
[5] Slabbinck, W., Dedecker, R., Vasireddy, S., Verborgh, R., Colpaert, P.: Linked Data Event Streams in Solid LDP containers. In: Proceedings of the 8th Workshop on Managing the Evolution and Preservation of the Data Web (MEPDaW) (2022)
[6] Vercruysse, A., Oo, S.M., Colpaert, P.: Describing a network of live datasets with the SDS vocabulary. In: Proceedings of the 8th Workshop on Managing the Evolution and Preservation of the Data Web (MEPDaW) (2022)

References

[7] S. Auer, L. Bühmann, C. Dirschl, O. Erling, M. Hausenblas, R. Isele, J. Lehmann, M. Martin, P. N. Mendes, B. Van Nuffelen, et al., Managing the life-cycle of linked data with the LOD2 stack, in: International semantic Web conference, Springer, 2012, pp. 1–16.

[8] K. M. Endris, S. Faisal, F. Orlandi, S. Auer, S. Scerri, Interest-based RDF update propagation, in: Proceedings of the 14th International Conference on The Semantic Web - ISWC 2015 - Volume 9366, Springer-Verlag, Berlin, Heidelberg, 2015, p. 513–529.

[9] M. Tasnim, D. Collarana, D. Graux, F. Orlandi, M.-E. Vidal, Summarizing entity temporal evolution in knowledge graphs, in: Companion Proceedings of The 2019 World Wide Web Conference, WWW '19, Association for Computing Machinery, New York, NY, USA, 2019, p. 961–965.

[10] A. Freitas, E. Curry, Big data curation, in: New Horizons for a Data-Driven Economy, 2016.

[11] J. Debattista, S. Auer, C. Lange, Luzzu—a methodology and framework for linked data quality assessment, J. Data and Information Quality 8 (2016).

[12] L. Gleim, S. Decker, Timestamped URLs as persistent identifiers, in: 6th Workshop on Managing the Evolution and Preservation of the Data Web (MEPDaW), 2020.

[13] J. D. Fernández, A. Polleres, J. Umbrich, Towards efficient archiving of dynamic linked open data, in: MEPDaW workshop at ESWC'15, 2015.

[14] O. Pelgrin, L. Galárraga, K. Hose, Towards fully-fledged archiving for RDF datasets, Semantic Web (2020) 1–24.