







#### 4.4.1 Library

The first dataset is a very small synthetic dataset about the availability of books in a library. The goal of this dataset is to explain the concept of changesets using books that are only available at specific moments in the library. In order to make it easily understandable, this dataset has only four versions, and 11 books that are available in specific versions of the dataset.

#### 4.4.2 DBpedia Live

A larger real-world dataset based on DBpedia Live [14] contains more than 48K unique triples over 89 versions. This dataset has been derived from the BEAR RDF archiving benchmark [9]. It contains the 100 most volatile resources from DBpedia Live over the course of three months with an hourly granularity.

## 5 CONCLUSIONS

RDF archiving has been an active area of research over the last couple of years. OSTRICH is storage and querying system for RDF archives that supports various kinds of versioned queries. With OSTRICH, versioned datasets can be stored efficiently, while at the same time enabling efficient support for versioned queries. When OSTRICH is combined with techniques such as Triple Pattern Fragments, versioned Linked Datasets can be published at a low cost, and complex SPARQL queries can be evaluated *in, between,* and *over* the different versions. This lowers the barrier towards historical analysis over datasets that evolve over time, such as biomedical patient information or certain taxonomies.

In the future, we will continue improving the performance of OSTRICH, and do an extensive performance evaluation. OSTRICH Admin will be kept up-to-date with the functionality of OSTRICH, so that OSTRICH datasets can be discovered and managed at a high-level using this Web application, without having to use the programmatic API for this.

## ACKNOWLEDGEMENTS

The described research activities were funded by Ghent University, imec, Flanders Innovation & Entrepreneurship (AIO), and the European Union. Ruben Verborgh is a postdoctoral fellow of the Research Foundation – Flanders.

## BIOGRAPHIES



**Ruben Taelman** is a PhD student at ID-Lab, Ghent University – imec, Belgium. His research concerns the server and client trade-offs for Linked Data publication and querying, with a particular focus on dynamic data, such as streams and versioning.



**Miel Vander Sande** is a post-doctoral researcher in Linked Data at Ghent University – imec. His main interest is low-cost Linked Data publishing infrastructures and Intelligent Web clients. In that regard, he executed numerous projects in Open Data legislation, digital publishing, e-learning, and data sharing.



**Ruben Verborgh** is a professor of Semantic Web technology at Ghent University – imec and a postdoctoral fellow of the Research Foundation Flanders. He explores the connection between Semantic Web technologies and the Web's architectural properties, with the ultimate goal of building more intelligent clients. Along the way, he became fascinated by Linked Data, REST/hypermedia, Web APIs, and related technologies.

## REFERENCES

- [1] Bizer, C., Heath, T., Berners-Lee, T.: Linked Data - the story so far. *Semantic Services, Interoperability and Web Applications: Emerging Concepts*. 205–227 (2009).
- [2] Umbrich, J., Decker, S., Hausenblas, M., Polleres, A., Hogan, A.: Towards dataset dynamics: Change frequency of Linked Open Data sources. *3rd International Workshop on Linked Data on the Web (LDOW)*. (2010).
- [3] Cyganiak, R., Wood, D., Lanthaler, M.: RDF 1.1: Concepts and Abstract Syntax. W3C, <http://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/> (2014).
- [4] Vander Sande, M., Colpaert, P., Verborgh, R., Coppens, S., Mannens, E., Walle, R. Van de: R&Wbase: git for triples. In: *Proceedings of the 6th Workshop on Linked Data on the Web (2013)*.
- [5] Neumann, T., Weikum, G.: x-RDF-3X: fast querying, high update rates, and consistency for RDF databases. *Proceedings of the VLDB Endowment*. 3, 256–263 (2010).
- [6] Volkel, M., Winkler, W., Sure, Y., Kruk, S.R., Synak, M.: Semversion: A versioning system for RDF and ontologies. In: *Second European Semantic Web Conference, ESWC 2005, Heraklion, Crete, Greece, May 29–June 1, 2005. Proceedings (2005)*.
- [7] Cerdeira-Pena, A., Farina, A., Fernández, J.D., Martínez-Prieto Miguel A.: Self-indexing RDF archives. In: *Data Compression Conference (DCC)*, 2016. pp. 526–535. IEEE (2016).
- [8] Taelman, R., Vander Sande, M., Verborgh, R., Mannens, E.: Versioned Triple Pattern Fragments: A Low-cost Linked Data Interface Feature for Web Archives. In: *Proceedings of the 3rd Workshop on Managing the Evolution and Preservation of the Data Web (2017)*.
- [9] Fernández, J.D., Umbrich, J., Polleres, A., Knuth, M.: Evaluating Query and Storage Strategies for RDF Archives. In: *Proceedings of the 12th International Conference on Semantic Systems. ACM (2016)*.
- [10] Fernández, J.D., Polleres, A., Umbrich, J.: Towards efficient archiving of Dynamic Linked Open Data. In: *Debattista, J., d'Aquin, M., and Lange, C. (eds.) Proceedings of the First DIACHRON Workshop on Managing the Evolution and Preservation of the Data Web*. pp. 34–49 (2015).
- [11] Meinhardt, P., Knuth, M., Sack, H.: TailR: a platform for preserving history on the web of data. In: *Proceedings of the 11th International Conference on Semantic Systems*. pp. 57–64. ACM (2015).
- [12] Verborgh, R., Vander Sande, M., Hartig, O., Van Herwegen, J., De Vocht, L., De Meester, B., Haesendonck, G., Colpaert, P.: Triple Pattern Fragments: a Low-cost Knowledge Graph Interface for the Web. *Journal of Web Semantics*. 37–38, (2016).
- [13] Fernández, J.D., Martínez-Prieto, M.A., Gutiérrez, C., Polleres, A., Arias, M.: Binary RDF Representation for Publication and Exchange (HDT). *Web Semantics: Science, Services and Agents on the World Wide Web*. 19, 22–41 (2013).
- [14] Morsey, M., Lehmann, J., Auer, S., Stadler, C., Hellmann, S.: DBpedia and the live extraction of structured data from wikipedia. *Program*. 46, 157–181 (2012).
- [15] McBride, B.: Jena: A semantic web toolkit. *IEEE Internet computing*. 6, 55–59 (2002).