



# Current research on theory and practice of digital libraries: best papers from TPDL 2017

Giannis Tsakonas<sup>1</sup> · Jaap Kamps<sup>2</sup>

Received: 3 April 2019 / Revised: 6 October 2019 / Accepted: 5 January 2020 / Published online: 14 February 2020  
© Springer-Verlag GmbH Germany, part of Springer Nature 2020

## Abstract

This volume presents a special issue on the 2017 edition of the theory and practice of digital libraries (TPDL) conference, held in Thessaloniki, Greece. We provide a brief overview of TPDL 2017 and introduce the selected papers that make up the rest of this volume. The papers cover different aspects of current digital library research, highlighting the important and multidisciplinary nature of the field.

**Keywords** Digital libraries · TPDL conference · Special issue

## 1 Introduction

For over 20 years, the International Conference on Theory and Practice of Digital Libraries (<https://www.tpd.eu/>), previously known as European Conference on Digital Libraries, constitutes one of the main discussion fora for matters of digital libraries. The 21st version of the International Conference on Theory and Practice of Digital Libraries took place in Thessaloniki, Greece during September 18–21, 2017. As any other event in this series of conferences, representatives from both fields of theory and practice cultivated an intense dialectic and shared results that form a state of the art imagery of digital libraries, that affect the life of academia, education, culture, society and industry.

The conference held the general theme “Part of the Machine: turning complex into scalable” and aimed to create a dialogue that would address the challenge of creatively transforming these highly synthesized environments into solutions that can scale for the benefit of varied communities. The general theme wanted to highlight the fact that digital libraries are complex systems that respond to the needs of multiple communities with escalating requirements.

Undoubtedly, the effect of big data in research and development is immense, and their collection, aggregation, analysis and interpretation is currently one of the main trends. However, smaller developments are equally interesting in giving the example of managing highly structured and organized information resources. The theme was divided into eight large areas or else tracks, namely Digital Humanities, E-Infrastructures, Information Retrieval, Semantics, Users and Societies, Content, Data and Services.

## 2 Revised and expanded papers

Out of the program of TPDL 2017, the General and the Program Committee Chairs nominated ten papers with the highest review scores. The authors of these papers were then invited to submit extended versions of their papers, something that materializes to—at least—30% new material. These extensions are needed not just to differentiate publications that a single research instance produces, but to give to the authors the opportunity to expand upon the description of their work and to provide more details on their technical approaches and results. Following the triple peer review process of the initial paper in the frame of the conference, a new review round started for the extended works by three reviewers, as the journal policy obliges. Finally, six papers were accepted for publication in this special issue.

“Tracking the History and Evolution of Entities: Entity-centric Temporal Analysis of Large Social Media Archives” by Fafalios et al. [2] is a paper that focuses on the analysis

---

✉ Jaap Kamps  
kamps@uva.nl

Giannis Tsakonas  
gtsak@upatras.gr

<sup>1</sup> University of Patras, Patras, Greece

<sup>2</sup> University of Amsterdam, Amsterdam, The Netherlands

of archived web resources, such as user generated content in social media and entities represented in user queries. The authors also define measures, such as entity popularity, attitude, controversiality and connectedness with other entities, in order to study how these were portrayed in social media in different time periods and under different aspects.

Freire et al. [3] in “Cultural heritage metadata aggregation using web technologies: IIIF, Sitemaps and Schema.org” explore some of the most recent web technologies and standards, such as the International Image Interoperability Framework (IIIF), Sitemaps and Schema.org, in the context of digital cultural heritage and especially in the frame of a large scale aggregation system, namely Europeana. The results of the authors’ exploration have been proved successful, especially taking into consideration the existing technology and the minimum acceptable implementation requirements in digital cultural heritage systems.

“Towards extracting event-centric collections from Web archives,” a paper authored by Gossen et al. [4], focuses on the development and testing of a classification method for web resources into event-centric collections. The authors treat this very intriguing issue for web archiving and retrieval by interlinking otherwise disconnected resources, based on the extraction of relevant information and on the relevance judgment in terms of topical and temporal proximity to user defined event-centric collections.

Pinto and Balke’s [6] “Assessing plausibility of scientific claims to support high-quality content in digital collections” was a very intriguing case of paper, as the authors use “claims” as key constructs of scientific papers to operationalize the notion of plausibility. This is a very ambiguous concept, and the investigation of its aspects in scholarly communication products aims at developing novel assessment techniques to support the work of reviewers. The authors experiment with neural embedding representations of text and topic models in papers from the PubMed digital library.

De Siqueira et al. [1] in “A pragmatic approach to hierarchical categorization of research expertise in the presence of scarce information” use supervised machine learning methods trained with manually labeled information to associate researchers and their expertise to a knowledge classification scheme. This classification is performed on the minimum information available, and it is possible to enhance organization processes in digital libraries and repositories, as well as in recommendation systems.

Walsh et al. [8] have authored “Characterising online museum users: a study of the National Museums Liverpool museum website,” a paper that looks into the information behavior of a large scale and wide coverage sample of visitors of an online museum. The authors were particularly concerned by the low levels of usage and bouncing rates in museum web presence systems and in their study worked

with a less convenient sample that would give them more information to explain this fact.

### 3 Envoi

The TPDL 2017 conferences also featured inspiring keynotes by Paul Groth (Elsevier/University of Amsterdam), Elton Barker (Pelagios/The Open University) and Dimitrios Tzovaras (CERTH/ITI). General information about the conference is available on its web pages [7], and all contributed papers and keynote abstracts are published in the conference proceedings [5]. Thessaloniki provided a very pleasant environment for intense discussion with colleagues and friends, with great food (and drinks), and surrounded by impressive Byzantine monuments. And many conference attendees managed to experience the time of Alexander the Great up-close-and-personal by visiting the tomb of his father King Philip II of Macedon in nearby Vergina. These served as vivid reminders of the important mission of the field to preserve, curate and give access to key information of the past and present in digital libraries.

As it is shown in the descriptions above, the papers in the rest of this volume cover a heterogeneous set of topics and highlight the diversity of the domain of digital libraries. It confirms one more that the components of this “machinery” are tightly interconnected and claim the skills and knowledge of multidisciplinary researchers. The editors of this special issue would like to thank the authors for the collaboration during the reparation of this issue and to extend their gratitude to the reviewers that shared their time with the community in order to provide constructive feedback to these seven papers. Their help, as well as the guidance by the Editorial team of IJDL, shows that this journal is a healthy community-driven publication that aims to offer deep, informative views on many aspects of digital libraries.

**Acknowledgements** Our main thanks are to the general chairs of TPDL’17 in Thessaloniki, in particular Yannis Manolopoulos, Lazaros S. Iliadis and Ioannis Karydis, to the IJDL editorial team, in particular Ingo Frommholz and Nabil Adam, and to the authors writing significantly extended versions of the earlier TPDL’17 papers.

**Funding** This funding was provided by NWO (Grant Nos. 314.99.108 and CISC.CC.016 )

### References

1. de Siqueira, G.O., Canuto, S., Gonçalves, M.A., Laender, A.H.F.: A pragmatic approach to hierarchical categorization of research expertise in the presence of scarce information. *Int. J. Digit. Libr.* **21** (2020). <https://doi.org/10.1007/s00799-018-0260-z>
2. Fafalios, P., Iosifidis, V., Stefanidis, K., Ntoutsis, E.: Tracking the history and evolution of entities: entity-centric temporal analysis of

- large social media archives. *Int. J. Digit. Libr.* **21** (2020). <https://doi.org/10.1007/s00799-018-0257-7>
3. Freire, N., Robson, G., Howard, J.B., Manguinhas, H., Isaac, A.: Cultural heritage metadata aggregation using web technologies: IIIF, sitemaps and schema.org. *Int. J. Digit. Libr.* **21** (2020). <https://doi.org/10.1007/s00799-018-0259-5>
  4. Gossen, G., Risse, T., Demidova, E.: Towards extracting event-centric collections from web archives. *Int. J. Digit. Libr.* **21** (2020). <https://doi.org/10.1007/s00799-018-0258-6>
  5. Kamps, J., Tsakonas, G., Manolopoulos, Y., Iliadis, L.S., Karydis, I. (eds.): Research and advanced technology for digital libraries. In: 21st International Conference on Theory and Practice of Digital Libraries, TPDL 2017, Thessaloniki, Greece, September 18–21, 2017, Proceedings, Lecture Notes in Computer Science, vol. 10450. Springer (2017). <https://doi.org/10.1007/978-3-319-67008-9>
  6. Pinto, J.M.G., Balke, W.T.: Assessing plausibility of scientific claims to support high-quality content in digital collections. *Int. J. Digit. Libr.* **21** (2020). <https://doi.org/10.1007/s00799-018-0256-8>
  7. TPDL'17: 21st International Conference on Theory and Practice of Digital Libraries (2017). <http://www.tpd1.eu/tpdl2017/>
  8. Walsh, D., Hall, M.M., Clough, P., Foster, J.: Characterising online museum users: a study of the national museums liverpool museum website. *Int. J. Digit. Libr.* **21** (2020). <https://doi.org/10.1007/s00799-018-0248-8>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.