

---

# Guest editorial: Digital language archives

Guest editorial

505

---

Digital language archives were most actively discussed during the early 2000s, with the Online Language Archives Community (OLAC) project funded by the U.S. National Science Foundation (NSF) leading these discussions, as well as the development of best practices and standards for these archives (Simons and Bird, 2003). After the end of the funding period, resources developed by OLAC continued to exist, yet there was a period when conversations about digital language archives were limited to the linguistics community (in particular, in response to the NSF mandate for all funded language revitalization research projects to deposit their data for reuse into digital archives). During this period, little attention was paid to this important area by the information science community, although digital language archives grew exponentially, with growth triggered by the NSF mandate and the development of initiatives, such as Digital Endangered Language and Music network and so on. In the recent years, the increasing interest in digital language archives in the information science community is evidenced by the IMLS-funded project investigating information organization in digital language archives (2018–2020) and several resulting publications and presentations (Burke *et al.*, 2021; Burke and Zavalina, 2019, 2020a, 2020b). It is also evidenced by the recent *IEEE/ACM Joint Conference on Digital Libraries Workshop* on digital language archives, which is intended to start a series of regular workshops held every other year and attracted presenters from six different countries from Asia, Australia, Europe, North and South America and Europe.

The goal of this Special Issue is to provide the first collection of publications that focus on various aspects of digital language archives theory and practice, history and directions. The Guest Editors sought submissions from information professionals and researchers who work on digital language archives problems, as well as from linguistics researchers, some of whom had rarely addressed the information science/practice audiences before, and *The Electronic Library* journal was a perfect platform for such interdisciplinary collaboration. Most of the papers included in this Special Issue are significantly extended versions of a selection of 14 short papers presented at the *First International Workshop on Digital Language Archives (LangArc '21)* held as part of the *Joint Conference on Digital Libraries* in September of 2021 (Zavalina and Chelliah, 2021).

The issue opens with the paper by well-known experts in the area of digital language archives – the linguists who led the OLAC project in the past. The papers in this issue represent a diverse range of topics of importance to digital language archives. Bird and Simons discuss what has been happening with OLAC resources and initiatives over the more than 20 years of its existence and present the new directions and opportunities for digital language archives in the future. The paper contributed by Weber discusses the general issues encountered with legacy language data in archives: those related to provenance, orphan data and citation tracking. The team led by Khait reports and discusses preliminary results of the large-scale international user study of language archives. Singh and colleagues discuss how they used a parametric approach in the analysis of platforms used for the development of digital language archives. Freitag discusses the issues that affect the development and functioning of the sociolinguistic archives in Brazil. Narayanan and Takhelambam discuss collaborative digital language archive development, using the case study of the Sikkim-Darjeeling Himalayas Endangered Language Archive (SiDHELA).



Two papers share the experiences of the Computational Resources of South Asian Languages (CoRSAL) digital language archive. One of them, contributed by Dale, presents the approaches tested by the CoRSAL in the development of their workflow for mediated archiving. In another paper, Burke and colleagues discuss the challenges and proposed solutions related to representing with metadata the names, subjects, and other important attributes of language data in the CoRSAL archive.

This Special Issue makes contributions to the field by sharing the best practices in digital language archives, including those related to making them more usable by underrepresented and underserved user communities (e.g. indigenous groups and speakers of endangered languages). It is not only the research and practical work of libraries, archives and museums that is affected by the growth of digital language archives, but also, importantly, education for linguists and information professionals. While this issue does not include articles focusing specifically on education, the Guest Editors believe research papers published in it will also be useful for informing training and curriculum development. They also anticipate that the Special Issue will help bring about changes in policy by drawing attention to the concerns related to improving access to digital language archives that support language revitalization, and to the need for funding initiatives to support addressing some of these issues.

**Oksana Zavalina**

*Department of Information Science, UNT, Denton, Texas, USA, and*

**Shobhana Lakshmi Chelliah**

*Department of Linguistics, University of North Texas, Denton, Texas, USA*

## References

- Burke, M. and Zavalina, O.L. (2019), "Exploration of information organization in language archives", *Proceedings of the Association for Information Science and Technology*, Vol. 56 No. 1, pp. 364-367, doi: [10.1002/pr2.30](https://doi.org/10.1002/pr2.30).
- Burke, M. and Zavalina, O.L. (2020a), "Descriptive richness of free-text metadata: a comparative analysis of three language archives", *Proceedings of the Association for Information Science and Technology*, Vol. 57 No. 1, p. e429, doi: [10.1002/pr2.429](https://doi.org/10.1002/pr2.429).
- Burke, M. and Zavalina, O.L. (2020b), "Identifying challenges for information organization in language archives", *iConference 2020: Sustainable Digital Communities*, pp. 622-629, doi: [10.1007/978-3-030-43687-2\\_52](https://doi.org/10.1007/978-3-030-43687-2_52).
- Burke, M., Zavalina, O.L., Phillips, M.E. and Chelliah, S. (2021), "Organization of knowledge and information in digital archives of language materials", *Journal of Library Metadata*, Vol. 20 No. 4, pp. 185-217.
- Simons, G. and Bird, S. (2003), "Building an open language archives community on the OAI foundation", *Library Hi Tech*, Vol. 21 No. 2, pp. 210-218.
- Zavalina, O.L. and Chelliah, S.L. (Eds) (2021), "Proceedings of the international workshop on digital language archives (LangArc '21)", available at: <https://doi.org/10.12794/langarc1851186>