Developing Graduate Curriculum for Digital Language Archive Stewardship

Oksana L. Zavalina[a] and Hugh J. Paterson III[b]

[a b] Department of Information Science, University of North Texas, USA

Oksana.Zavalina@unt.edu, i@hp3.me

ABSTRACT

Our interdisciplinary team including information and language scholars and educators and practicing information professionals at University of North Texas (UNT) and Indiana University (IU) is developing an evidence-based practice-oriented online curriculum to train library, archives, and museums (LAM) professionals in the archiving, curation, and ethical dissemination of resources that provide the means to revitalize community memory and language. During this 2-year project, over 30 LAM students will complete the project-developed UNT graduate course INFO 5385 Community Language Archiving and Curation for Information Professionals, and additional estimated 100 students at UNT and IU will complete individual project-developed learning modules integrated into other courses. Development of the learning materials is informed by digital language archive practices, research, and existing training materials for archive depositors. Resulting open-access adaptable learning resources are expected to be widely used by academic programs and educators for training LAM students and by participatory archives as continuing education resources.

ALISE RESEARCH TAXONOMY TOPICS

Education of information professionals -- Curriculum; Information services -- Archives; Information services -- Community engagement; Information practices -- Specific populations; Information practices -- Information needs.

AUTHOR KEYWORDS

community language archives; ethical stewardship; knowledge dissemination; digital curation; language documentation.

Copyright 2024 by the authors. Published under a Creative Commons Attribution-ShareAlike 4.0 International License. See https://creativecommons.org/licenses/by/4.0.

DOI: https://doi.org/10.21900/j.alise.2024.1657

INTRODUCTION AND BACKGROUND

The emergence of digital workflows in linguistics, anthropology, and other related disciplines has resulted in increased need for digital preservation of language-focused cultural heritage artifacts (Berez, 2015; Boerger et al., 2023). Language resources are often created in the context of scholarly or academic collaborations with ethnolinguistic minority communities (Bischoff & Jany, 2018; Czaykowska-Higgins, 2009; Miyashita & Many Bears, 2009; Rice, 2011). They also often have specific and technical audiences (Holton, 2012; Woodbury, 2014). These materials not only represent the truths discovered through scientific inquiry but also have ethical, cultural, legal, and social dimensions which must be addressed by memory institutions for the well-being of the past, present, and future of stakeholder communities and their linguistic cultural heritage (Kendall, 2014; Seyfeddinipur et al., 2019).

Within scholarly communities there is an increased requirement for reproducibility (Baker, 2016; National Academies of Sciences, 2019; Oreskes et al., 1994; Srivastava, 2018). Also, empirical evidence-informed science—including language sciences—requires that evidence be examinable (Berez-Kroeker et al., 2018; Draxler & Trouvain, 2020; Garellek et al., 2020). Depositing research project data to language archives is also motivated by funding requirements, and sometimes scholarly work is funded by governments with public access mandates (e.g., Nelson, 2022). Memory institutions and scholarly repositories play a significant role in furthering social evolution and innovation by stewarding these resources.

Within ethnolinguistic communities there is a growing interest in documenting their cultural and linguistic practices which are losing diversity due to globalization, urbanization, industrialization (Conathan, 2011; Bromham et al., 2021). Often communities are turning to commercial social media solutions rather than dedicated memory institutions (Christen, 2019; Ungsitipoonporn et al., 2021). These practices highlight the need for language archives to have direct lines of communication with the communities whose cultural heritage they steward. There is a gap in the graduate-level training preparing LAM professionals to support these organization-to-community relationships.

THE NEED

Language archives (independent organizations or special collections within other organizations) are administrative architectures for managing language resources. Often an archive stewards a variety of collections containing resources on or about one or more minority languages. Since language is a crucial part of cultural heritage, language resources often require special administrative considerations due to sensitive political contexts and histories of oppression. Some communities are very open about sharing their cultural and linguistic heritage, while others are more protective. Language archives traditionally serve several audiences including language scholars, language community members, educators, and language advocates. Audience considerations impact accessibility, discovery, and the types of collections held at memory institutions. Additionally, collections may have very different arrangements, inclusion criteria, and curation needs depending on their creation context.

The need for English-language formal graduate-level training on stewardship issues specifically relevant to language resource collections is currently not met by the ALA-accredited degree programs. Several stewardship organizations have created relevant but tangential continuing education resources for depositors (mainly language scholars). These include: *The Endangered Languages Archive's Annual Training for Grantees of Endangered Languages Documentation Programme,*[1] *The Pacific and Regional Archive for Digital Sources in Endangered Cultures's Guidelines for Depositors* (Miller, 2023), and The US National Science Foundation funded *Archiving for the Future*, an online course (Kung et al., 2020).[2] Another training (also targeted for language scholars) is Dallas International University's graduate-level curriculum for linguistics students focusing on language documentation (Boerger et al., 2023). Tailored to the audience of language community members, the *Collaborative Digital Language Archiving Curriculum*[3] aims to guide language users through the deposit process (CORSAL, n.d).

Language archives have a growing social role. This includes supporting the development of language-education resources. However, none of the existing language-archive trainings specifically can serve as curricular materials for LAM professionals tasked with stewarding language resources. Filling this gap is the motivation for our current efforts. Following the ideas outlined by Zavalina and Chelliah (2021), we present ongoing work conducted with federal funding to develop graduate-level curriculum for LAM professionals who are interested in working with collections of language resources and collaborating with stakeholders of these collections. Our curriculum is informed by participatory approaches to stewardship and engagement with memory institutions. Below we discuss the focus and content of the project-developed training materials, and specific types of roles filled by information professionals that are the target audiences for this training.

CURRICULUM OVERVIEW

Our curriculum development is underpinned by a user centric paradigm and informed by digital language archive practices and existing training materials for archive depositors (critically evaluated by our team in Coronado and Zavalina, 2024). We rely on constructivist theory (Swan, 2005) for student application of knowledge and incorporate action learning theory (Waddill, 2006) into the learning exercises. The course's introductory nature is reflected in its depth of coverage, topics, and in the skills required of students. We anticipate that this will ease the adoption of the open-access course into a broader set of institutional curricula.

The following outline presents the course learning objectives.

1. Module 1: Planning, developing and managing a community language archive
    a. Define community language archives and their functions.

[1] https://www.eldp.net/en/our+trainings/resource
[2] https://archivingforthefuture.teachable.com
[3] https://corsal.unt.edu/curriculum

    b. Describe procedures and considerations for community language archive planning and development.
2. Module 2: Ethical archival practices and digital curation for community language archives
    a. Examine archival theory and practice and digital curation trends and perspectives relevant for community language archives.
    b. Explain important access, preservation, and description issues and practical problems associated with community language archives
3. Module 3: Metadata, digital content management, and web archiving for community language archives
    a. Identify metadata standards that can be utilized in representing community language archive materials to support information needs and user tasks of community language archive users.
    b. Describe the problems related to selection of digital content management tools, depositing, and web archiving for community language archives.
4. Module 4: Dissemination, use, and evaluation of community language archives
    a. Identify efficient and ethical ways of disseminating the content of community language archives.
    b. Discuss approaches for the evaluation of the services provided by community language archives.

Below, we highlight some of the topics covered in these materials that—based on the available relevant literature and practical evidence—deserve special attention during training of LAM professionals charged with language-archive stewardship. This includes *Fiduciary responsibility, Language identification, Resource relationships, Resource of-ness, Interactive modality and materiality***,** and the *Evaluation of community language archives.* The selection of these areas of emphasis is informed by assessments of students' and practitioners' knowledge gaps (e.g., Aljalahmah & Zavalina, 2023a, 2023b; Zavalin, 2023; Zavalin & Zavalina, 2023; Zavalina & Burke, 2021).

Fiduciary responsibility. Agreements, donor expectations, and legal requirements are some of the thorniest and most conflict-inducing areas of stewardship. The legal concept of fiduciary responsibility can help guide discussions to ensure that stewarding organizations are free from future encumberment and maintain positive social presence, while also guiding resource-management activities.

Language identification. Identification and representation of a resource's languages is a complex and sometimes confusing issue that has not been an easy task for LAM professionals in the past as documented for example in the BABEL report (PCC, 2020). Our curriculum informs students of various approaches for representing speech varieties and train students to work with different controlled vocabularies for language and dialect representations.

Resource relationships. Language resources are rarely created in isolation and, therefore, have a high density of relationships to other resources. Understanding these relationships and representing them in metadata is important for discovery. LAM professionals need to know how to use bibliographic models to analyze relations between language resources, including

Work-Expression-Manifestation-Item (WEMI) entities first defined in the Functional Requirements for Bibliographic Records (IFLA, 1998; Riva et al., 2017; Coyle, 2022). The WEMI entities representation is highly relevant for language resources due to their complex structures including the part-whole relationships, which has been a challenge for curators to analyze (e.g., Zavalina & Burke, 2021). Also, a persistent issue across language archives is the lack of following the one-to-one principle, according to which a metadata record should represent a single resource (Sugimoto et al., 2002). Language archives often subsume several different works and/or expressions under a single metadata record. This can lead to metadata-management and resource-management complications, particularly regarding rights and access (e.g., Huber, 2023; Paterson III, 2021). Our curriculum addresses specifics of representing complex relationships between language resources.

Resource of-ness, interactive modality and materiality. Of-ness representation facilitates discovery for multiple audiences through describing what a resource is of. Users of language archives often start their search for resources with of-ness in mind, thinking in terms like: 'of-a-language' or 'of-a-family-member' (e.g., Anderson, 2005; Kipp, 2007; Thieberger & Harris, 2022). Identifying of-ness is normally a part of the subject analysis and representation (e.g., Hjørland, 2016; Yee, 1990). Our curriculum, following recent discussions related to research datasets (Friedrich & Siegers, 2016; O'Dell, 2016), emphasizes representing of-ness in description of language resources (tabular, textual, visual, and aural), thereby supporting their discovery. In addition, information resources have interactive modality (i.e., the ways that one can engage with the artifact: e.g., via aural or visual methods, via reading, etc.). They also have interactive materiality (i.e., available ways to integrate an artifact's content into life activities: e.g., is it only viewable online, is it intended for integration in other processes, does one download it for use via mobile devices, etc.). Interactive modality and interactive materiality of the resources are represented with various controlled vocabularies. Our course introduces students to these, with the focus on their usefulness for community language archive users.

Evaluating community language archives. Evaluation of information repositories of any kind is important for improving user experiences. Assessing the extent to which a repository or collection meets user tasks, first defined in FRBR model (IFLA, 1998), is one of the ways to evaluate bibliographic databases that are intended to facilitate discovery, including discovery of language archive resources (e.g., Zavalin, 2023). Evaluations of language archives must consider the needs and tasks of two very different major groups of end-users: language community members and academic audiences. As the project-developed learning materials center on community language archives, they are informed by what our team has learned from language communities about their information access needs and expectations, i.e., how well they are met in the digital representation and interactions with language resources. Our curriculum introduces LAM professionals to several perspectives from which evaluation of these archives can be conducted. Evaluated areas include services, technical platforms, content description and arrangement, as well as discovery.

WHO NEEDS THIS TRAINING?

While various topics in stewardship of information resources in general are covered by archival academic programs, as stated above, specifics of stewarding language resources are not represented in this training, and the project-developed learning materials are intended to fill this important gap. Information professionals in several key roles—which in the United States are found commonly in academic libraries—need exposure to issues in the management of language resources. The roles broadly construed are:

- Informational professionals responsible for inventory and implementation of data management plans
- Informational professionals working with university faculty on adding legacy-data collections to repositories
- Informational professionals whose responsibilities include establishing a new archive and/or development of existing digital collections
- Programming Librarians / Exhibit Production Specialists
- Curators[4] and Collection Maintenance Professionals
- Informational professionals working on system/repository migration tasks.

The functions and related processes—of importance to language archives— that LAM professionals in these roles are engaged in include the following:

1. **Review of prospective deposits including data management plans.** Prior to submission, many grant applications must undergo internal review. Recently, the United States funding agencies started requiring data retention and public access to all publicly funded research outputs (Nelson, 2022). As data retention is often a function of the university library, it is expected that academic library information professionals will be increasingly reviewing data management plans for language documentation projects conducted by linguists (and other language-related projects conducted by anthropologists and other researchers) for any requirements upon and associated costs incurred in support of language resource stewardship and access.

2. **Programming / Exhibit Production.** Programming Librarians are called upon to produce exhibits—temporary online interactive experiences containing source materials and the derivative scholarly record (Ferreira et al., 2021; Paterson III, 2023). Increasingly, exhibits are funded for short periods of time through libraries. Two notable examples are the British Library's *True Echoes*[5] project and the Hugh Broady collection at the University of Cape Town (Jones & Muftic, 2020). In this capacity, the information professional is central to branding decisions and crafting the exhibit's materiality, balancing the interests of the research participants and co-creators (e.g., underrepresented or endangered language communities), scholars, and the stewardship organization.

3. **Digital Repository set-up.** The information professional may be called upon to set up the digital infrastructure needed for long-term stewardship of language-resource collections. In this context the professional needs to select and customize the appropriate software

[4] We include the tasks of acquisition and accession of new and legacy collections under the role of curator. Digital and analog collections might be managed very differently from each other.
[5] https://www.true-echoes.com

tools for the types of language resources and structure of language collections stewarded, while keeping in mind the institution's mission and long-term goals. Memory institutions have differing social contexts, and how they engage users with resources varies by institutional mission (Dietz et al., 2005; Gilliland-Swetland, 2000; Paterson III, 2023). This can have a significant impact on interactions of end-users with digital collections.

4. **Curation, Collection Maintenance & Migration.** Curators and collection-maintenance professionals may undertake tasks which include digitization, file conversions (Harvey et al., 2015), metadata standardization (Lynch et al., 2020), and updating metadata when standards change (Zavalina et al., 2016; Zavalina & Zavalin, 2019); or they may engage with audiences to enhance descriptions of resources following initial accession (Budzise‐Weaver et al., 2012; Van Hyning & Jones, 2021). Crucial for accessibility is how metadata in a collection of language materials fits with other memory institution practices. Understanding when to apply isolating approaches to maintain metadata integrity is important. For example, the metadata scheme used in a digital language collection may be more detailed than metadata schemas commonly used across memory institutions. Information system migration may necessitate conversion of metadata to a less-detailed standard, which results in information loss (Zinn et al., 2016; Trippel & Zinn, 2018). Standard controlled vocabularies are subject to editing policies (e.g., Zavalina & Zavalin, 2017). For example, with changing the creator or contributor name in the language archive metadata to match the changed heading (authorized form of the personal name) in the authority record of the LC/NACO name authority file, the creation context of the archival collection may be lost. It may be necessary to set up a local (language-collection-specific) name authority file to prevent such loss, as well as to provide normalized forms of names for those community members serving in the roles of resource creators and contributors who are not represented in standard ones.

The growing prominence of community language-archive materials in digital repositories hosted by memory institutions highlights the need for updating and expanding the training for information professionals engaged in the functions discussed above. The curriculum developed by our team aims to address this need.

CONCLUSION

Our project is committed to publishing an open access curriculum which will equip students with knowledge and skills for the ethical and effective stewardship of language resources at memory institutions. The course has been in development since 2023. The first cohort of students took it in the summer of 2024, with continued curricular refinement through 2025. To maximize this curriculum's value, the team is getting feedback from underserved language community representatives and LAM and language documentation experts. Our project seeks to position language archives as crucial stakeholders serving language communities in ongoing access to their cultural heritage. The training gives information professionals tools to directly engage with communities in the archival stewardship of language resources to their benefit, while also supporting the needs of other user groups (including researchers). Resulting open-access adaptable learning resources will be made openly available for reuse by LAM degree programs in formal education and by archives as continuing education resources.

ACKNOWLEDGEMENTS

This project is supported by the Institute of Museum and Library Services Laura Bush 21st Century Librarian program (grant RE-254860-OLS-23). The authors of this paper acknowledge curriculum development contributions of the project team members: Dr. Shobhana Chelliah, Dr. Ana Roeschley, Dr. Brian O'Connor, Dr. Mark Phillips, Sergio Coronado, and Merrion Frederick.

REFERENCES

Aljalahmah, S., & Zavalina, O. L. (2023a). Exploration of metadata practices in digital collections of archives with Arabian language materials. In O. L. Zavalina & S. L. Chelliah (Eds.), *Proceedings of the International Workshop on Digital Language Archives: LangArc-2023* (pp. 11–14). University of North Texas. https://doi.org/10.12794/langarc2114295

Aljalahmah, S., & Zavalina, O. L. (2023b). Student-created Dublin Core metadata representing Arabic language eBooks: Comparison of individual and group work outcomes. *Journal of Education for Library and Information Science*, e20230016. https://doi.org/10.3138/jelis-2023-0016

Anderson, J. (2005). Access and control of indigenous knowledge in libraries and archives: Ownership and future use. American Library Association and The MacArthur Foundation Columbia University. https://static1.squarespace.com/static/5a1c710fbce17620f861bf47/t/5b249d0ff950b7fedff16822/1529126160548/paper_anderson.pdf

Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature*, 533(7604), 452–454. https://doi.org/10.1038/533452a

Berez, A. (2015). Integrating archiving into the language documentation curriculum at the University of Hawai'i at Mānoa. In A. Harris, N. Thieberger, & L. Barwick (Eds.), *Research, Records and Responsibility: Ten Years of PARADISEC* (pp. 39–52). Sydney University Press. http://www.jstor.org/stable/j.ctt1bh4b3f.5

Berez-Kroeker, A. L., Gawne, L., Kung, S. S., Kelly, B. F., Heston, T., Holton, G., Pulsifer, P., Beaver, D. I., Chelliah, S., Dubinsky, S., Meier, R. P., Thieberger, N., Rice, K., & Woodbury, A. C. (2018). Reproducible research in linguistics: A position statement on data citation and attribution in our field. *Linguistics*, 56(1), 1–18. https://doi.org/10.1515/ling-2017-0032

Bischoff, S. T., & Jany, C. (Eds.). (2018). Insights from practices in community-based research: From theory to practice around the globe. De Gruyter. https://doi.org/10.1515/9783110527018

Boerger, B. H., Moeller, S. R., Reiman, W., & Self, S. (2023). Language and Culture Documentation Manual. LeanPub. https://leanpub.com/languageandculturedocumentationmanual

Bromham, L., Dinnage, R., Skirgård, H., Ritchie, A., Cardillo, M., Meakins, F., Greenhill, S., & Hua, X. (2021). Global predictors of language endangerment and the future of linguistic diversity. *Nature Ecology & Evolution*, 6(2), 163–173. https://doi.org/10.1038/s41559-021-01604-y

Budzise‐Weaver, T., Chen, J., & Mitchell, M. (2012). Collaboration and crowdsourcing: The cases of multilingual digital libraries. *The Electronic Library*, 30(2), 220–232. https://doi.org/10.1108/02640471211221340

Christen, K. (2019). "The songline is alive in Mukurtu": Return, reuse, and respect. In L. Barwick, J. Green, & P. Vaarzon-Morel (Eds.), Archival returns: Central Australia and beyond (pp. 153–172). University of Hawai'i Press & Sydney University Press. http://hdl.handle.net/10125/24882

Computational Resource for South Asian Languages (CoRSAL). (n.d.). Collaborative Digital Language Archiving Curriculum. University of North Texas. https://corsal.unt.edu/curriculum

Conathan, L. (2011). Archiving and language documentation. In J. Sallabank & P. K. Austin (Eds.), *The Cambridge Handbook of Endangered Languages* (pp. 235–254). Cambridge University Press. https://doi.org/10.1017/CBO9780511975981.012

Coronado, S. I., & Zavalina, O.L., & (2024). Digital language archiving: Reuse and adaptation of non-LIS learning materials in LIS education. In, *The Ethics and Evolution of Truth and Information: Proceedings of the Association for Library and Information Science Education Annual Conference: ALISE 2024* (9 pp.). https://doi.org/10.21900/j.alise.2024.1736

Coyle, K. (2022). Works, expressions, manifestations, items: An ontology. *The Code4Lib Journal*, 53. https://journal.code4lib.org/articles/16491

Czaykowska-Higgins, E. (2009). Research models, community engagement, and linguistic fieldwork: Reflections on working within Canadian indigenous communities. *Language Documentation & Conservation,* 3(1), 15–50. https://hdl.handle.net/10125/4423

Dietz, S., Besser, H., Borda, A., Geber, K., & Lévy, P. (2005). Virtual museum (of Canada): The next generation. Canadian Heritage Information Network. https://www.academia.edu/34788702/Steve_Dietz_Howard_Besser_Ann_Borda_and_Kati_Geber_eds_2005_Virtual_Museum_of_Canada_the_Next_Generation_Canadian_Heritage_Information_Network

Draxler, C., & Trouvain, J. (2020). On principles of phonetic archiving: From paleo-phonetics to modern speech data management. *In, Proceedings of the 3rd Workshop on History of Speech Communication Research* (pp. 41–47). https://www.coli.uni-saarland.de/~trouvain/Draxler_Trouvain_2019.pdf

Ferreira, V., Lukschy, L., Watyam, B., Ungsitipoonpor, S., & Seyfeddinipur, M. (2021). A website is a website is a website: Why trusted repositories are needed more than ever. In O. L. Zavalina & S. L. Chelliah (Eds.), *Proceedings of the International Workshop on Digital Language Archives: LangArc 2021* (pp. 1–4). University of North Texas. https://doi.org/10.12794/langarc1851176

Friedrich, T., & Siegers, P. (2016). The ofness and aboutness of survey data: Improved indexing of social science questionnaires. In A. F. X. Wilhelm & H. A. Kestler (Eds.), *Analysis of Large and Complex Data* (pp. 629–638). Springer International Publishing. https://doi.org/10.1007/978-3-319-25226-1_54

Garellek, M., Simpson, A., Roettger, T. B., Recasens, D., Niebuhr, O., Mooshammer, C., Michaud, A., Lee, W.-S., Kirby, J., Gordon, M., & Yu, K. M. (2020). Letter to the editor: Toward open data policies in phonetics: what we can gain and how we can avoid pitfalls. *Journal of Speech Sciences*, 9, 3–16. https://doi.org/10.20396/joss.v9i00.14955

Gilliland-Swetland, A. J. (2000). The societal role of archives. In *Enduring Paradigm, New Opportunities: The Value of the Archival Perspective in the Digital Environment* (pp. 4–6). Council on Library and Information Resources. https://www.clir.org/wp-content/uploads/sites/6/pub89.pdf

Harvey, M. J., Mason, N. J., McLean, A., Murray-Rust, P., Rzepa, H. S., & Stewart, J. J. P. (2015). Standards-based curation of a decade-old digital repository dataset of molecular information. *Journal of Cheminformatics*, 7(1), Article: 43. https://doi.org/10.1186/s13321-015-0093-3

Hjørland, B. (2016). Subject (of documents). In B. Hjørland & C. Gnoli (Eds.), *ISKO Encyclopedia of Knowledge Organization*. International Society for Knowledge Organization. http://www.isko.org/cyclo/subject

Huber, C. (2023). Why it can be difficult to make historic language recordings accessible: A View from a corpus of historic dialect recordings. In O. L. Zavalina & S. L. Chelliah (Eds.), *Proceedings of the International Workshop on Digital Language Archives: LangArc-2023* (pp. 15–18). University of North Texas. https://doi.org/10.12794/langarc2114302

IFLA Study Group on the Functional Requirements for Bibliographic Records & Plassard, Marie-France. (1998). Functional Requirements for Bibliographic Records: Final Report (19; 2nd ed., [UBCIM Publications, New Series] IFLA Series on Bibliographic Control). K.G. Saur. https://www.ifla.org/wp-content/uploads/2019/05/assets/cataloguing/frbr/frbr.pdf

Jones, K., & Muftic, S. (2020). Endangered African languages featured in a digital collection: The case of the ǂKhomani San | Hugh Brody Collection. In R. Mabuya, P. Ramukhadi, M. Setaka, V. Wagner, & M. van Zaanen (Eds.), *Proceedings of the First workshop on Resources for African Indigenous Languages (RAIL): Language Resources and Evaluation Conference (LREC 2020)*, Marseille, 11–16 May 2020 (pp. 1–8). European Language Resources Association (ELRA). https://www.aclweb.org/anthology/2020.rail-1.1

Kendall, T. (2014). Archiving and managing sociolinguistic data: The problems of portability, access and security, and discoverability and relevance. *Language and Linguistics Compass*, 8(11), 495–504. https://doi.org/10.1111/lnc3.12108

Kipp, D. (2007). Swimming in words. *Cultural Survival Quarterly*, 31(2), 36–43. https://www.culturalsurvival.org/publications/cultural-survival-quarterly/swimming-words

Kung, S. S., Sullivant, R., Pojman, E., & Niwagaba, A. (2020). Archiving for the future: Simple steps for archiving language documentation collections. Teach Online with Teach:able. https://archivingforthefuture.teachable.com

Lynch, J. D., Gibson, J., & Han, M.-J. (2020). Analyzing and normalizing type metadata for a large aggregated digital library. *The Code4Lib Journal*, 47. https://journal.code4lib.org/articles/14995

Miller, J. C. (2023). PARADISEC Workflows. PARADISEC. https://paradisec-archive.github.io/PARADISEC_workflows

Miyashita, M., & Many Bears, R. (2009). Academic-community collaboration: Data collection, linguistics research, and language teaching. *44th International Conference for Salish and Neighboring Languages*. https://lingpapers.sites.olt.ubc.ca/files/2018/02/2009_Miyashita_ManyBears.pdf

National Academies of Sciences. (2019). Replicability. In *Reproducibility and Replicability in Science*. National Academies Press (US). https://doi.org/10.17226/25303

Nelson, A. (2022). Memorandum for the heads of executive departments and agencies: Ensuring Free, immediate, and equitable access to federally funded research [Memo]. United States. Office of Science and Technology Policy; United States. Executive Office of the President. https://doi.org/10.21949/1528361

O'Dell, A. J. (2016). Defining usefulness and facilitating access based on research applications. *IFLA-WLIC: Satellite meeting - Subject Access: Unlimited Opportunities*, Columbus Ohio. https://library.ifla.org/id/eprint/2042/1/S12-2016-odell-en.pdf

Oreskes, N., Shrader-Frechette, K., & Belitz, K. (1994). Verification, validation, and confirmation of numerical models in the earth sciences. *Science, New Series*, 263(5147), 641–646. http://www.jstor.org/stable/2883078

Paterson III, H. J. (2021). Audio artifacts in language archives: From cataloguing to referencing and citation [Presentation Slides; Video]. *2021 Conference of the International Association of Sound and Audiovisual Archives: Closing the Gap for a New Generation of Sound and Audiovisual Archives*, Hosted online. https://hughandbecky.us/Hugh-CV/talk/2021-audio-artifacts-in-language-archives/

Paterson III, H. J. (2023). Diversity and identity: Categories for OAI data-providers in the Open Language Archives Network. *Proceedings from North American Symposium on Knowledge Organization (NASKO)*, 9(1), 18–31. https://doi.org/10.7152/nasko.v9i1.16301

Program for Cooperative Cataloging (PCC), Standing Committee on Standards (SCS), BIBFRAME and MARC Bibliographic Encoding for Languages (BABEL) Task Group. (2020). BABEL Final Report. US Library of Congress. https://www.loc.gov/aba/pcc/taskgroup/BABEL-Final-Report.docx

Rice, K. (2011). Documentary linguistics and community relations. *Language Documentation & Conservation*, 5, 187–207. http://hdl.handle.net/10125/4498

Riva, P., Le Bœuf, P., & Žumer, M. (Eds.). (2017). IFLA Library Reference Model: A Conceptual Model for Bibliographic Information (December 2017). International Federation of Library Associations and Institutions (IFLA). https://www.ifla.org/publications/node/11412

Seyfeddinipur, M., Ameka, F., Bolton, L., Blumtritt, J., Carpenter, B., Cruz, H., Drude, S., Epps, P. L., Ferreira, V., Galucio, A. V., Hellwig, B., Hinte, O., Holton, G., Jung, D., Buddeberg, I. K., Krifka, M., Kung, S., Monroig, M., Neba, A. N., … Woodbury, T. (2019). Public access to research data in language documentation: Challenges and possible strategies. *Language Documentation & Conservation*, 13, 545–563. http://hdl.handle.net/10125/24901

Srivastava, S. (2018). Verifiability is a core principle of science. *Behavioral and Brain Sciences*, 41, e150. https://doi.org/10.1017/S0140525X18000869

Sugimoto, S., Baker, T., & Weibel, S. L. (2002). Dublin Core: Process and principles. In E.-P. Lim, S. Foo, C. Khoo, H. Chen, E. Fox, S. Urs, & T. Costantino (Eds.), *Digital Libraries: People, Knowledge, and Technology* (Vol. 2555, pp. 25–35). Springer Berlin Heidelberg. https://doi.org/10.1007/3-540-36227-4_3

Swan, K. (2005). A constructivist model for thinking about learning online. In J. C. Moore & J. Bourne (Eds.), *Elements of Quality Online Education: Engaging Communities* (pp. 13–30). The Sloan Consortium.

Thieberger, N., & Harris, A. (2022). When your data is my grandparents singing: Digitisation and access for cultural records, the Pacific and Regional Archive for Digital Sources in Endangered Cultures (PARADISEC). *Data Science Journal*, 21, 9. https://doi.org/10.5334/dsj-2022-009

Trippel, T., & Zinn, C. (2018). Lessons learned: On the challenges of migrating a research data repository from a research institution to a university library. In N. Calzolari, K. Choukri, C. Cieri, T. DeClerck, S. Goggi, K. Hasida, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, S. Piperidis, & T. Tokunaga (Eds.), *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)* (pp. 146–150). European Language Resources Association (ELRA). https://aclanthology.org/L18-1023

Ungsitipoonporn, S., Buachut, W., Ferreira, V., & Seyfeddinipur, M. (2021). Community archiving of ethnic groups in Thailand. *Language Documentation & Conservation*, 15, 267–284. http://hdl.handle.net/10125/24975

Van Hyning, V. A., & Jones, M. A. (2021). Data's destinations: Three case studies in crowdsourced transcription data management and dissemination. *Startwords*, 2. https://doi.org/10.5281/zenodo.5750691

Waddill, D. D. (2006). Action e-learning: An exploratory case study of action learning applied online. *Human Resource Development International*, 9(2), 157–171. https://doi.org/10.1080/13678860600616347

Yee, M. M. (1990). Subject access to moving image materials in a MARC-based online environment. *Beyond the Book*, 1(1), 97–116. https://escholarship.org/uc/item/1nz230kw

Zavalin, V. I. (2023). Ukrainian archival metadata in WorldCat: Exploratory analysis. In O. L. Zavalina & S. L. Chelliah (Eds.), *Proceedings of the International Workshop on Digital Language Archives: LangArc-2023* (pp. 38–41). University of North Texas. https://doi.org/10.12794/langarc2114298

Zavalin, V., & Zavalina, O. L. (2023). Exploration of accuracy, completeness and consistency in metadata for physical objects in museum collections. In I. Sserwanga, A. Goulding, H. Moulaison-Sandy, J. T. Du, A. L. Soares, V. Hessami, & R. D. Frank (Eds.), *Information for a Better World: Normality, Virtuality, Physicality, Inclusivity* (pp. 83–90). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-28032-0_7

Zavalina, O. L., & Burke, M. (2021). Assessing skill building in metadata instruction: Quality Evaluation of Dublin Core metadata records created by graduate students. *Journal of Education for Library and Information Science*, 62(4), 423–442. https://doi.org/10.3138/jelis.62-4-2020-0083

Zavalina, O. L., & Chelliah, S. (2021). Exploring language archiving education for information professionals and interdisciplinary collaboration to support information access. ALISE 2021 Conference Proceedings. *ALISE 2021 Proceedings: Crafting a Resilient Future: Leadership, Education, & Inspiration*, September 20-24, 2021, Virtual. IDEALS. https://hdl.handle.net/2142/110944

Zavalina, O. L., & Zavalin, V. (2017). Identity management analysis: An empirical investigation into the state of library community's authority data conformance to the new standard. *Knowledge Discovery and Data Design Innovation*, 233–248. https://doi.org/10.1142/9789813234482_0012

Zavalina, O. L., & Zavalin, V. (2019). Evaluation of metadata change in authority data over time: An effect of a standard evolution. *Proceedings of the Association for Information Science and Technology*, 55(1), 593-597. https://doi.org/10.1002/pra2.2018.14505501064

Zavalina, O. L., Zavalin, V., Shakeri, S., & Kizhakkethil, P. (2016). Developing an empirically-based framework of metadata change and exploring relation between metadata change and metadata quality in MARC library metadata. *Procedia Computer Science*, 99, 50–63. https://doi.org/10.1016/j.procs.2016.09.100

Zinn, C., Trippel, T., Kaminski, S., & Dima, E. (2016). Crosswalking from CMDI to Dublin Core and MARC 21. In N. Calzolari, K. Choukri, T. Declerck, S. Goggi, M. Grobelnik, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, & S. Piperidis (Eds.), *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)* (pp. 2489–2495). European Language Resources Association (ELRA). https://aclanthology.org/L16-1395