From CV to OLAC

Hugh Paterson III

Unaffiliated Collaborator

i@hp3.me

The Open Language Archive Community (OLAC, Simons & Bird 2003) metadata standards are foundational in sharing the existence of language materials. OLAC was designed to promote resource discoverability across institutional archives. However, anyone can tell OLAC where to look for language materials by publishing a feed (a string of XML data) and pointing OLAC to the feed.

In 2012 at the *Satellite Workshop for Sociolinguistic Archive Preparation* (LSA, Portland) it was expressed that many people, who might have data, do not have the desire to formally archive their data based on institutional policies. They do not want to openly share the data itself, but they do want people to know it exists. Even if scholars do archive in their institutional repositories, broad-scope search engines (GoogleScholar) may choose to ignore these resources Arlitsch and O'Brien (2012). Academically-oriented social networks (academia.edu and researchgate.net) attempt to meet this market demand for discoverability (Ovadia 2014; Niyazov, et al. (2016). These solutions fail to aggregate resources based on field-specific formal ontologies. As a result, the work of many retired and late-career linguists is being "lost" to convenient platforms. Convenience also drives some early-career linguists to use costless-deposit archives (Zenodo and OSF) without considering discoverability.

OLAC lists 390,035 items across participating archives (as of July 2020); Glottolog lists 338,158 items as of version 4.2.1. However these numbers do not necessarily reflect the actual language resources archived. Several archives have not recently updated their metadata feed. For instance, the SIL Language and Culture Archive reports that it has 46,000[1] items (August 2020) but only reports 30,177 to OLAC (last updated in 2013). Even if language materials are deposited in an OLAC participating archive, they may not be discoverable through OLAC. By our calculations, the number of linguistic descriptions

---

[1] https://web.archive.org/web/20200816131036/https://www.sil.org/resources/language-culture-archives

grows by approximately 15,000 items per year. Many of these items never find their way into OLAC or Glottolog.

We hold that the best way for researchers to optimize their social profile — advertising their experience and academic output — is to archive their content within institutional repositories but self-host a CV-oriented website, pushing metadata to aggregators. Using open source technologies following strategies outlined by Utomo & Falahah (2020) we use Hugo[2] and the WowChemy[3] theme. With a modified RSS feed that compiles custom metadata and produces an OLAC compliant data feed, the researcher can then advertise language descriptions through OLAC.

Suggested Citation

Paterson III, Hugh J. 2021. From CV to OLAC. Paper presented at: The 7th International Conference on Language Documentation & Conservation (ICLDC), 4–7 March. University of Hawai'i at Mānoa.

References

Arlitsch, Kenning and Patrick S. O'Brien. 2012. Invisible institutional repositories: Addressing the low indexing ratios of IRs in Google Scholar. *Library Hi Tech* 30(1). 60-81. https://www.emerald.com/insight/content/doi/10.1108/07378831211213210/full/html (2020-05-09 17:10:04.) doi:10.1108/07378831211213210

Niyazov, Yuri, Carl Vogel, Richard Price, Ben Lund, David Judd, Adnan Akil, Michael Mortonson, Josh Schwartzman & Max Shron. 2016. Open Access Meets Discoverability: Citations to Articles Posted to Academia.edu. *PLOS ONE* 11(2). e0148257. https://dx.plos.org/10.1371/journal.pone.0148257 (2020-09-30 18:54:43.) doi:10.1371/journal.pone.0148257

Ovadia, Steven. 2014. ResearchGate and Academia.edu: Academic Social Networks. *Behavioral & Social Sciences Librarian* 33(3). 165-69. http://www.tandfonline.com/doi/abs/10.1080/01639269.2014.934093 (2020-09-30 18:55:06.) doi:10.1080/01639269.2014.934093

Simons, Gary F. & Steven Bird. 2003. The Open Language Archives Community: An Infrastructure for Distributed Archiving of Language Resources. *Literary and Linguistic Computing* 18(2). 117-28. doi:10.1093/llc/18.2.117

Utomo, Prayudi & Falahah. 2020. Building Serverless Website on GitHub Pages. *IOP Conference Series: Materials Science and Engineering* 879. 012077. https://iopscience.iop.org/article/10.1088/1757-899X/879/1/012077 (2020-09-30 18:43:57.) doi:10.1088/1757-899X/879/1/012077

---

[2] https://gohugo.io
[3] https://wowchemy.com