# Archives: features without a business plan - Handout

Language archives effectively achieve their stated goal of long-term data storage, but they often fail to meet the needs of both community member users and also researcher users. We present the mismatch between the services offered by 'language archives' and the jobs to be done by both segments of archive service consumers as a business model mismatch. We present this mismatch as a business risk which mitigates against the mission of archives. Finally, we discuss some technical innovations which would bring workflow revisions and ease consumer friction, but we hold that this is ultimately a business model problem and not a technological problem, because an organization's business plan is what drives it choice of technological deployment.

Some Basic Terms

Data storage
A general term for the storing of data in electromagnetic or other forms for use by a computer or device.[1]
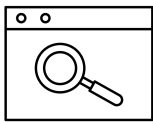
Access point ("External View")
Data can be centrally managed, but power different websites and applications. These independent websites and external views we call *Access Points*. Access points are usually designed for meeting specific end-user needs (based on end-user "jobs-to-be-done").

Producer / Submitter
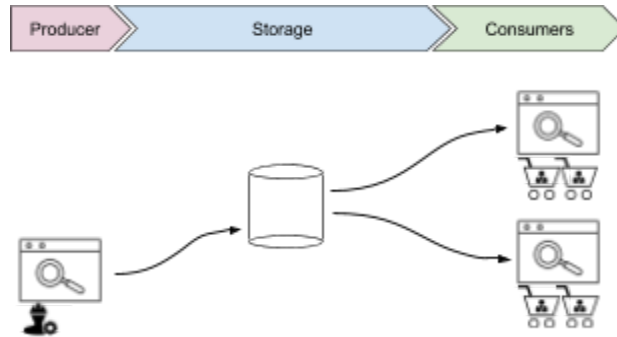Some who submits, modifies, and accesses stored data.

Consumer / User
Someone who accesses stored data but does not submit or modify any of that data in its original storage location.
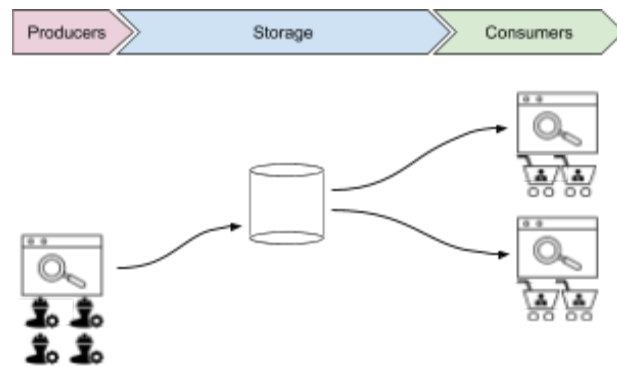

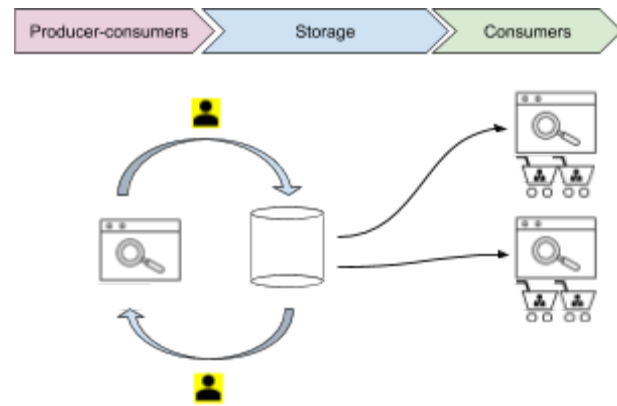[1] https://www.techopedia.com/definition/23342/data-storage

Content interaction models

The Lone-wolf Traditional Model

The Multi-Producer Model

The Prosumer Model

Special Containers

Archive
Examples: Alaska Native Language Center[2]; Brazil's National Museum[34]
Mission: to preserve the informational content, the medium, and the media; often mediums beyond digital only exist in these sorts of institutions.

Data Repository
Examples: University of North Texas[5]; Dataverse[6]
Mission: A place to start the data acquisition for a new project, Where can Researchers dump their materials? — frequently sizable data sets, but usually not "big data".

Institutional Repository
Examples: UO Scholars' Bank[7], SIL Language and Culture Archive[8]; ScholarlyCommons @ UPenn[9]
Mission: Preserve the documents of the organization for the benefit of the organization[10].

Project Specific Repository
Example: Hausar Baka[11], Russell Schuh Northern Nigerian Language Materials [12] Roger Blench website[13], ASIS, Atlante Sintattico dell'Italia Settentrionale[14] Recorded Hausa Materials Archive[15]
Mission: Broadcast informational data/code sharing by Lone-wolf self-publishing, small organization, departmental research lab, departmental grant funded project on departmental website.

---

[2] https://www.uaf.edu/anlc/mission
[3] Consider also the: Natural History Museum in New Delhi and the Instituto Butantan in São Paulo.
[4] https://www.wired.com/story/brazil-museum-fire-digital-archives
[5] https://digital.library.unt.edu/explore/collections/UNTDRD
[6] https://dataverse.org
[7] https://scholarsbank.uoregon.edu
[8] https://www.sil.org/resources/language-culture-archives
[9] https://repository.upenn.edu
[10] Some Organizations such as BePress (subsidiary of Elsevier) mix terms and suggest that people institutional faculty should "archive" at institutional repositories. See BePress subscriber only presentation: "Why Archive in an Institutional Repository, Marianne Buehler, University of Nevada Las Vegas".
[11] https://web.archive.org/web/20171005100105/http://aflang.linguistics.ucla.edu/hausarbaka
[12] https://www.nsf.gov/awardsearch/showAward?AWD_ID=0111289
[13] https://www.rogerblench.info
[14] http://asis-cnr.unipd.it
[15] http://www.iu.edu/~celtie/hausa_archive.html

Parts of a collection

Accession: A group of things which arrives at a processing facility at the same time

Archive: Some people call a curated set of documents an archive - this gets confused with "an archival institution". And also confused with the software and set of servers holding our digital data.

Collection: Institutions usually break portions of their things (in our case data) into "collections" so that they can be more effectively managed. An institutional archive might have several dozen collections. New virtual collections may be created pulling and mixing content for new audiences.

Items: (*Collections* - but hopefully not), *Documents*, *Books*, *Papers*, *Media*, *Manuscripts*, *Corpra*.

Corpus: A curated, balanced (for some explicit purpose) set of digital artifacts (generally unpublished content pieces), which are treated as a single unit for some research purpose.
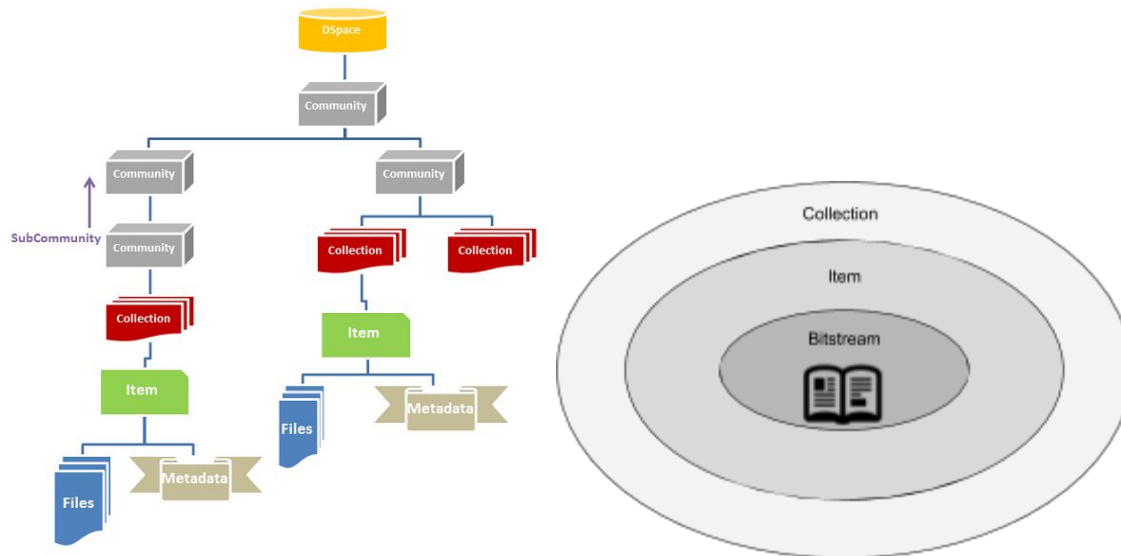
Jobs to be done

These are the jobs which require us to share data, and the methods that we use to share data. What are our tools of choice to accomplish the jobs (tasks) which requires sharing?

In linguistics and language documentation we all do tasks; some are repetitive, and some are repeated in every new language documentation project. Frequently, tasks involve the application of methods and increasingly require us to collaborate. Can you think of three procedures or methods which you have used in your linguistics work?

1. _____
2. _____
3. _____

Many times we are also simultaneously collaborating with various people in various places. Can you name three things you have had to share in the process of a professional collaboration?

4. _____
5. _____
6. _____

What was the primary factor which caused you to choose that tool?

7. _____
_____
_____

See the questionnaire handout!

Risk analysis

Within the academy we have long held that our work should have value. But what value an academic's work should have, or what qualifies as "valuable" and "valuable to whom" has been tenuous. By putting our work in depositories and archives we are transferring value in new ways, and we are engaging in new economies of value exchange. With any new economic venture there are inherent risks. We use the risk-analysis tool below to develop hypotheses about what kinds of tools will allow us to best leverage new economies of value exchange.

	Social Benefit	
	+	-
Single-platform Archive as Social Sole Solution (Single point of engagement) +	• • • • •	• • • • •
Multi-platform Archive as Long-term storage with targeted interactions offloaded to other interaction platforms -	• • • • •	• • • • •

We posit the following statements as hypotheses related to past and current assessments of the value of an academic's work:

1. In the past, academics were measured on how many quantifiable "widgets" they produced. (i.e. successful Ph.D Dissertations chaired, grants awarded, or academic papers published)
   - There has been change over the last 20 years to move the assessment from how many widgets are produced to how much value is produced by the individual for the institution or department.
2. Publications that bring influence are how we traditionally show that we've brought value to our community.
   - Altmetrics suggests that we can capture a metric for that value when we have citations or "back links". However, part of bringing value to a community includes bringing behavioral change. There is currently no metric for demonstrating that a publication has created behavioral change, in how we do our business of linguistics and language documentation.

Image Credits

Town Hall by Creaticca Creative Agency from the Noun Project
Repository by Nithinan Tatah from the Noun Project
Office building by Gregor Cresnar from the Noun Project
Repository by lastspark from the Noun Project
Website by ibrandify from the Noun Project
Web Page by Mat fine from the Noun Project
File by Markus from the Noun Project
Archives by apukspaks from the Noun Project
Book by Yo! Baba from the Noun Project
DSpace Item Structure:
https://wiki.duraspace.org/display/DSDOC6x/Functional+Overview#FunctionalOverview-Onlineaccesstoyourdigitalassets

Consulted Literature

Bow, Catherine, Michael Christie & Brian Devlin. 2014. Developing a Living Archive of
Aboriginal Languages. *Language Documentation & Conservation* 8. 345–360.
http://scholarspace.manoa.hawaii.edu/handle/10125/24612 (20 April, 2019).

Bel, Bernard & Médéric Gasquet-Cyrus. 2014. Digital curation and event-driven methods at
the service of endangered languages. In Mari C. Jones (ed.), *Endangered Languages
and New Technologies*, 113–126. Cambridge: Cambridge University Press.
doi:10.1017/CBO9781107279063.011.
http://ebooks.cambridge.org/ref/id/CBO9781107279063A018 (25 April, 2019).

Bird, Steven & Gary F. Simons. 2000a. A Survey of the State of the Art in Digital Language
Documentation and Description. Digital. Online, ms.
http://www.language-archives.org/documents/survey.html (23 April, 2019).

Bird, Steven & Gary F. Simons. 2000b. White Paper on Establishing an Infrastructure for
Open Language Archiving. Digital. Online, ms.
http://www.language-archives.org/documents/white-paper.html (23 April, 2019).

Collister, Lauren B. & Ashley L. Taylor. 2017. Altmetrics, Legacy Scholarship, and Scholarly
Legacy. *Pennsylvania Libraries: Research & Practice* 5(2). 127–137.
doi:10.5195/PALRAP.2017.154. http://palrap.org/ojs/index.php/palrap/article/view/154
(9 September, 2018).

Duffy, Brooke Erin & Jefferson D. Pooley. 2017. "Facebook for Academics": The
Convergence of Self-Branding and Social Media Logic on Academia.edu. *Social Media
+ Society* 3(1). 205630511769652. doi:10.1177/2056305117696523.
http://journals.sagepub.com/doi/10.1177/2056305117696523 (26 April, 2019).

Dobrin, Lise M. & Saul Schwartz. 2016. Collaboration or Participant Observation? Rethinking Models of "Linguistic Social Work." 10. 253–277. http://scholarspace.manoa.hawaii.edu/handle/10125/24694 (20 April, 2019).

Gawne, Lauren & Andrea L. Berez-Kroeker. 2018. Reflections on Language Documentation 20 Years after Himmelmann 1998. In Bradley McDonnell, Andrea L. Berez-Kroeker & Gary Holton (eds.), *Reflections on reproducible research* (Language Documentation & Conservation Special Publication 15), 22–32. Honolulu, Hawai'i: University of Hawai'i Press. http://scholarspace.manoa.hawaii.edu/handle/10125/24805 (20 April, 2019).

Glenn, Akiemi. 2009. Five Dimensions of Collaboration: Toward a Critical Theory of Coordination and Interoperability in Language Documentation. *Language Documentation & Conservation* 3(2): 149-160. http://scholarspace.manoa.hawaii.edu/handle/10125/4437 (20 April, 2019).

Hanke, Florian R. 2017. Computer Supported Collaborative Language Documentation. Melbourne, Australia: The University of Melbourne Ph.D Dissertation. http://hdl.handle.net/11343/192578.

Heesen, Remco 2018. Why the Reward Structure of Science Makes Reproducibility Problems Inevitable: *The Journal of Philosophy* 115(12). 661–674. doi:10.5840/jphil20181151239. http://www.pdcnet.org/oom/service?url_ver=Z39.88-2004&rft_val_fmt=&rft.imuse_id=jphil_2018_0115_0012_0661_0674&svc_id=info:www.pdcnet.org/collection (26 April, 2019).

Holton, Gary. 2011. "Unknown Unknowns" and the Retrieval Problem in Language Documentation and Archiving. *Language Documentation & Conservation* 5. 157–168. http://scholarspace.manoa.hawaii.edu/handle/10125/4496 (20 April, 2019).

Holton, Gary. 2012. Language archives: They're not just for linguists any more. In Frank Seifart, Geoffrey Haig, Nikolaus P. Himmelmann, Dagmar Jung, Anna Margetts & Paul Trilsbeek (eds.), *Potentials of Language Documentation: Methods, Analyses, and Utilization* (Language Documentation & Conservation Special Publication 3), 111–117. Honolulu, Hawai'i: University of Hawai'i Press. http://scholarspace.manoa.hawaii.edu/handle/10125/4523 (20 April, 2019).

Holton, Gary, Kavon Hooshiar & Nick Thieberger. 2017. Developing collection management tools to create more robust and reliable linguistic data. *Proceedings of the 2nd Workshop on the Use of Computational Methods in the Study of Endangered Languages*, 33–38. Honolulu, Hawai'i: Association for Computational Linguistics. doi:10.18653/v1/W17-0105. http://aclweb.org/anthology/W17-0105 (24 April, 2019).

Johnson, Heidi. 2004. Language Documentation and Archiving, or How to build a better Corpus. In Austin, Peter (ed.), *Language Documentation and Description* (2), 140-53. London, UK: SOAS. http://www.elpublishing.org/PID/026

Klamer, M, Trilsbeek, P, Hoogervorst, T and Haskett, C. 2017. Creating a Language Archive of Insular South East Asia and West New Guinea. In: Odijk, J and van Hessen, A. (eds.) *CLARIN in the Low Countries*, Pp. 113–121. London: Ubiquity Press. DOI: https://doi.org/10.5334/bbi.10

Leonard, Wesley Y. & Erin Haynes. 2010. Making "collaboration" collaborative: An examination of perspectives that frame linguistic field research. *Language Documentation & Conservation* 4. 269–293. http://scholarspace.manoa.hawaii.edu/handle/10125/4482 (20 April, 2019).

Mamtora, Jayshree & Catherine Bow. 2017. Towards a Unique Archive of Aboriginal Languages: A Collaborative Project. *Journal of the Australian Library and Information Association* 66(1). 28–41. doi:10.1080/00049670.2017.1282845. https://www.tandfonline.com/doi/full/10.1080/00049670.2017.1282845 (20 April, 2019).

Meakins, Felicity, Jennifer Green & Myfany Turpin. 2018. *Understanding linguistic fieldwork* (Understanding Language Series 1). First edition. London : New York : Routledge/ Taylor & Francis Group.

Nathan, David. 2013. Access and Accessibility at ELAR, A Social Networking Archive for Endangered Languages Documentation. In Mark Turin & Claire Wheeler (eds.), *Oral Literature in the Digital Age: Archiving Orality and Connecting with Communities* (World Oral Literature Series 2). Open Book Publishers. doi:10.11647/OBP.0032.03. http://www.openbookpublishers.com/product/186/ (24 April, 2019).

Nordmoe, Jeremy. 2011. Introducing RAMP: an application for packaging metadata and resources offline for submission to an institutional repository. In Proceedings of Workshop on Language Documentation & Archiving 18 November 2011 at SOAS, London. Edited by: David Nathan. p. 27-32. http://www.sil.org/acpub/repository/LDLT3_Nordmoe_preprint.pdf

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. (Ed.) Pablo Dorta-González. *PLOS ONE* 11(2). e0148257. doi:10.1371/journal.pone.0148257. https://dx.plos.org/10.1371/journal.pone.0148257 (26 April, 2019).

Osterwalder, Alexander. 2004. *The business model ontology a proposition in a design science approach*. Lausanne, Switzerland: Université de Lausanne Ph.D Dissertation. http://doc.rero.ch/record/4210.

Osterwalder, Alexander, Yves Pigneur, Gregory Bernarda, Alan Smith & Trish Papadakos. 2015. *Value Proposition Design : How to Create Products and Services Customers Want.* Somerset, New Jersey: Wiley.

Osterwalder, Alexander, Yves Pigneur & Tim Clark. 2010. *Business model generation: a handbook for visionaries, game changers, and challengers*. Hoboken, New Jersey: Wiley.

Paterson, Hugh J. III & Jeremy Nordmoe. 2013. *Challenges of implementing a tool to extract metadata from linguists: the use case of RAMP*. Poster presented at 3rd International Conference on Language Documentation and Conservation, at the University of Hawai'i Mānoa, Honolulu, HI. February 28 - March 3rd.

Paterson, Hugh J. III. 2015. *Lexical dataset archiving: an assessment of practice*. Poster presented at the 4th International Conference on Language Documentation and Conservation, at the Ala Moana Hotel in Honolulu, HI. February 26th – March 1st Version 1.0

Robinson, Laura C. 2006. Archiving Directly from the Field. In Linda Barwick & Nick Thieberger (eds.), Sustainable data from digital fieldwork: proceedings of the conference held at the University of Sydney, 4-6 December 2006, 23–32. Sydney: Sydney University Press. https://ses.library.usyd.edu.au/bitstream/2123/1291/1/4FH2006.pdf.

Shepard, Michael Alvarez. 2016. The Value-Added Language Archive: Increasing Cultural Compatibility for Native American Communities. 10. 458–479. http://scholarspace.manoa.hawaii.edu/handle/10125/24715 (20 April, 2019).

Simons, Gary F. & Steven Bird. 2000a. Requirements on the Infrastructure for Open Language Archiving. Digital. Online, ms. http://www.language-archives.org/documents/requirements.html (23 April, 2019).

Simons, Gary F. & Steven Bird. 2000b. The Seven Pillars of Open Language Archiving: A Vision Statement. Digital. Online, ms. http://www.language-archives.org/documents/vision.html (23 April, 2019).

Watt, Adrienne. 2014. Data Modelling. *Database Design*. 2nd-version 1.2 edn. Victoria, British Columbia, Canada: BCcampus. https://opentextbc.ca/dbdesign01/chapter/chapter-5-data-modelling/ (25 April, 2019).

Wasson, Christina, Gary Holton & Heather S. Roth. 2016. Bringing User-Centered Design to the Field of Language Archives. 641–681. http://scholarspace.manoa.hawaii.edu/handle/10125/24721 (20 April, 2019).

Wasson, Christina, Melanie Medina, Miyoung Chong, Brittany LeMay, Emma Nalin & Kenneth Saintonge. 2018. Designing for Diverse User Groups: Case Study of a

Language Archive. *Journal of Business Anthropology* 7(2). 235.
doi:10.22439/jba.v7i2.5605. https://rauli.cbs.dk/index.php/jba/article/view/5605 (25
April, 2019).

Xie, Iris & Krystyna K. Matusiak. 2016. User needs and search behaviors. *Discover Digital Libraries*, 231–253. Elsevier. doi:10.1016/B978-0-12-417112-1.00008-9.
https://linkinghub.elsevier.com/retrieve/pii/B9780124171121000089 (26 April, 2019).

Zhang, Allison B. & Don Gourley. 2009. Digital collections management system. *Creating Digital Collections*, 125–148. Elsevier. doi:10.1016/B978-1-84334-396-7.50009-2.
https://linkinghub.elsevier.com/retrieve/pii/B9781843343967500092 (26 April, 2019).