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\(^3\)
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\(^4\); Dataverse\(^5\)
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](https://www.uaf.edu/anlc/mission)
\(^3\) Consider also the: Natural History Museum in New Delhi and the Instituto Butantan in São Paulo.
\(^5\) [https://digital.library.unt.edu/explore/collections/UNTDRD](https://digital.library.unt.edu/explore/collections/UNTDRD)
\(^6\) [https://dataverse.org](https://dataverse.org)
\(^7\) [https://scholarsbank.uoregon.edu](https://scholarsbank.uoregon.edu)
\(^8\) [https://www.sil.org/resources/language-culture-archives](https://www.sil.org/resources/language-culture-archives)
\(^9\) [https://repository.upenn.edu](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”.
\(^13\) [https://www.rogerblench.info](https://www.rogerblench.info)
\(^14\) [http://asis-cnr.unipd.it](http://asis-cnr.unipd.it)
\(^15\) [http://www.iu.edu/~celtie/hausa_archive.html](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.

<table>
<thead>
<tr>
<th></th>
<th>Social Benefit</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>+</td>
</tr>
<tr>
<td><strong>Single-platform</strong></td>
<td></td>
</tr>
<tr>
<td>Archive as Social Sole Solution (Single point of engagement)</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Multi-platform</strong></td>
<td></td>
</tr>
<tr>
<td>Archive as Long-term storage with targeted interactions offloaded to other interaction platforms</td>
<td>-</td>
</tr>
</tbody>
</table>

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.
Collaborators: Hugh Paterson III and Richard Griscom
Language Description & Revitalization Workgroup at the University of Oregon

Image Credits

Town Hall by Creaticca Creative Agency from the Noun Project
Repository by Nithinan Tatah from the Noun Project
Office building by Gregor Crennar 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

http://scholarspace.manoa.hawaii.edu/handle/10125/24612 (20 April, 2019).

doi:10.1017/CBO9781107279063.011.


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