

## Abstract

This poster documents the design and development of a language archive for indigenous communities in North-east India that speak Tibeto-Burman languages. Although a number of existing archives include data on some of these languages, the materials are not available in formats that can easily be used for linguistic and cultural revitalization. A core part of the archive is participatory research with members of the communities. This ensures that the contents of archive, and the different experiences enabled by the content, accommodate local cultural practices of information access, sharing and use.

*“The language itself shows our identity and the richness of community is all expressed by language. Whether it is a folk song, folk tales or in riddles or proverbs or whatever that we use it all rests through the medium of language. So, it is an important thing. And without the language we are nothing.”*

– Sumshot Khular, member of the Lamkang Community



## Digital language archives of yesterday

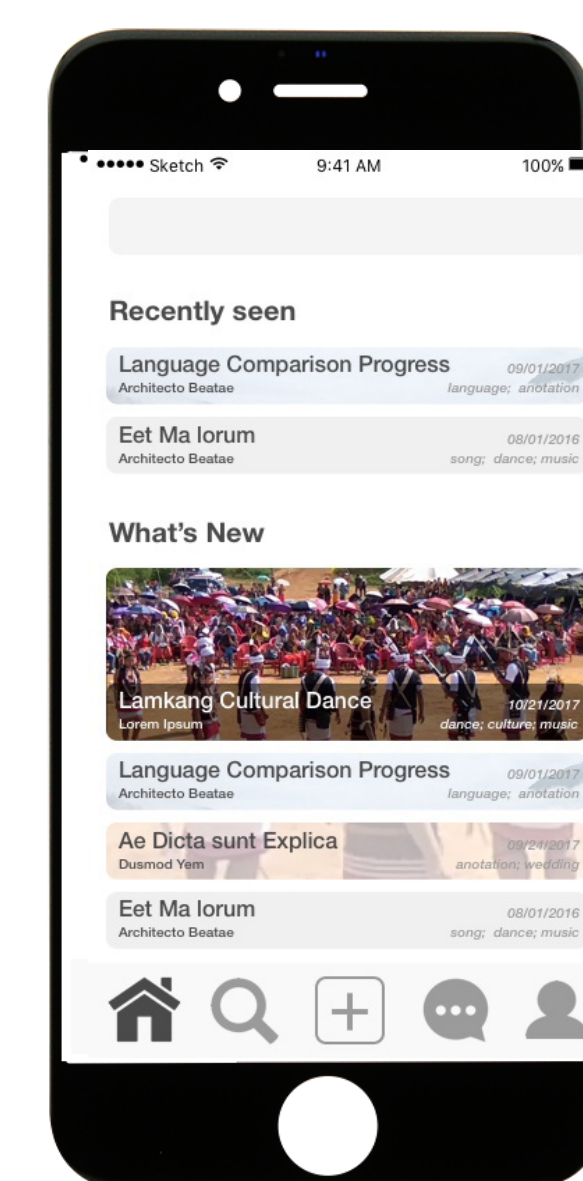
- Video and audio recordings showcasing the genius of a language
- Grammar descriptions
- Available for future generations
- Safe, ethical, shareable
- Annotated by professionals
- Metadata for searchability
- Data ingestion by archivist

## Exploratory User Research for Computational Resource for South Asian Languages: CoRSAL

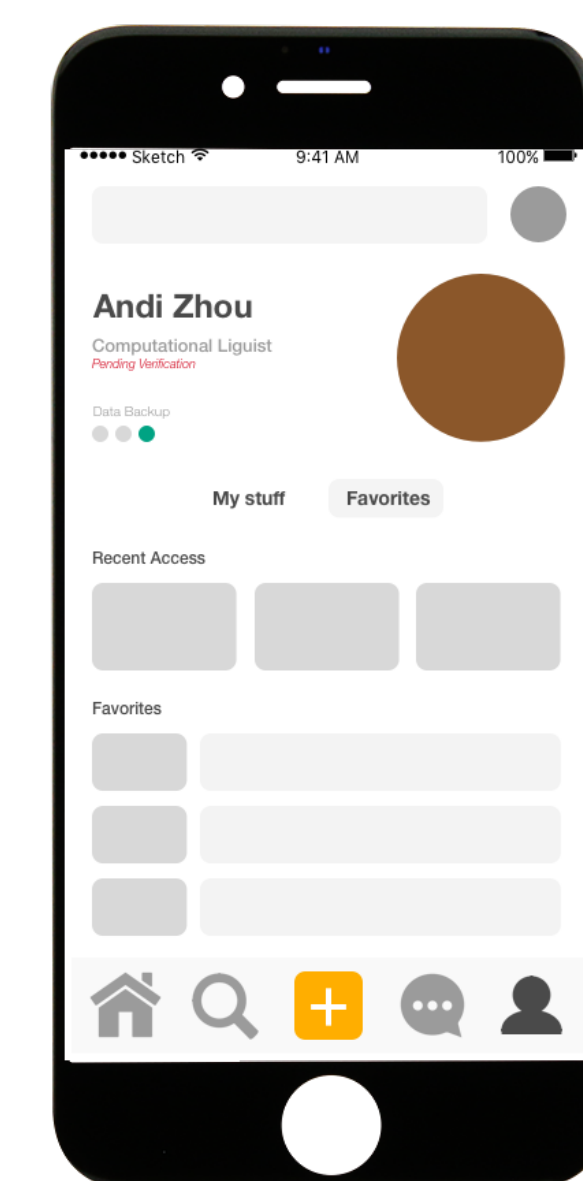
In order to make CoRSAL as useful as possible for all of the intended user groups, Christina Wasson’s Design Anthropology class at the University of North Texas conducted research in Fall 2016 with four user groups: language communities, linguists, computational linguists, and depositors. The students collected data on barriers to language archive use and developed recommendations for CoRSAL to meet the user groups’ diverse needs. Results are summarized in the table:

	Language Communities	Linguists	Computational Linguists	Depositors
<b>Group Description</b>	<ul style="list-style-type: none"> <li>• Speakers of languages whose materials are deposited in language archive</li> <li>• CoRSAL’s focus is on Tibeto-Burman languages</li> </ul>	<ul style="list-style-type: none"> <li>• Typically main user group considered</li> <li>• Rarely use archives as source of research data</li> </ul>	<ul style="list-style-type: none"> <li>• Relatively new field</li> <li>• Create speech/text processing system or human machine translation and interaction systems</li> </ul>	<ul style="list-style-type: none"> <li>• Deposit linguistic data into language archives</li> <li>• Multi-step process of preparing linguistic materials for deposit</li> </ul>
<b>Barriers</b>	<ul style="list-style-type: none"> <li>• Lack of standardized orthography</li> <li>• Lack of language learning materials</li> <li>• Inconsistent Internet and access to computers</li> </ul>	<ul style="list-style-type: none"> <li>• Data difficult to find and access</li> <li>• Lack of standards for metadata and annotation</li> </ul>	<ul style="list-style-type: none"> <li>• Inconsistencies among file formats and annotation styles</li> <li>• Availability of relevant data</li> </ul>	<ul style="list-style-type: none"> <li>• Time-consuming process of deposit preparation</li> <li>• Lack of professional recognition for deposit prep</li> <li>• Difficulty of updating deposited materials</li> <li>• Colonial legacy issues</li> </ul>
<b>Design Implications and Recommendations</b>	<ul style="list-style-type: none"> <li>• Encourage active role of community members in archive development</li> <li>• Support literacy efforts in language community</li> <li>• Make hard copies of materials available</li> <li>• Develop mobile application for language archive</li> </ul>	<ul style="list-style-type: none"> <li>• Improve data accessibility through interface design and search</li> <li>• Develop thoughtful standards for metadata</li> <li>• Encourage best practices for annotation</li> </ul>	<ul style="list-style-type: none"> <li>• Avoid use of certain file types</li> <li>• Encourage depositors to provide annotation information</li> <li>• Allow for downloads of data subsets</li> <li>• Make data machine-readable</li> </ul>	<ul style="list-style-type: none"> <li>• Streamline language archive depositing process</li> <li>• Encourage universities to recognize deposits as publications</li> <li>• Enable depositors to easily update their materials</li> <li>• Specify usage rights</li> <li>• Empower language communities</li> </ul>

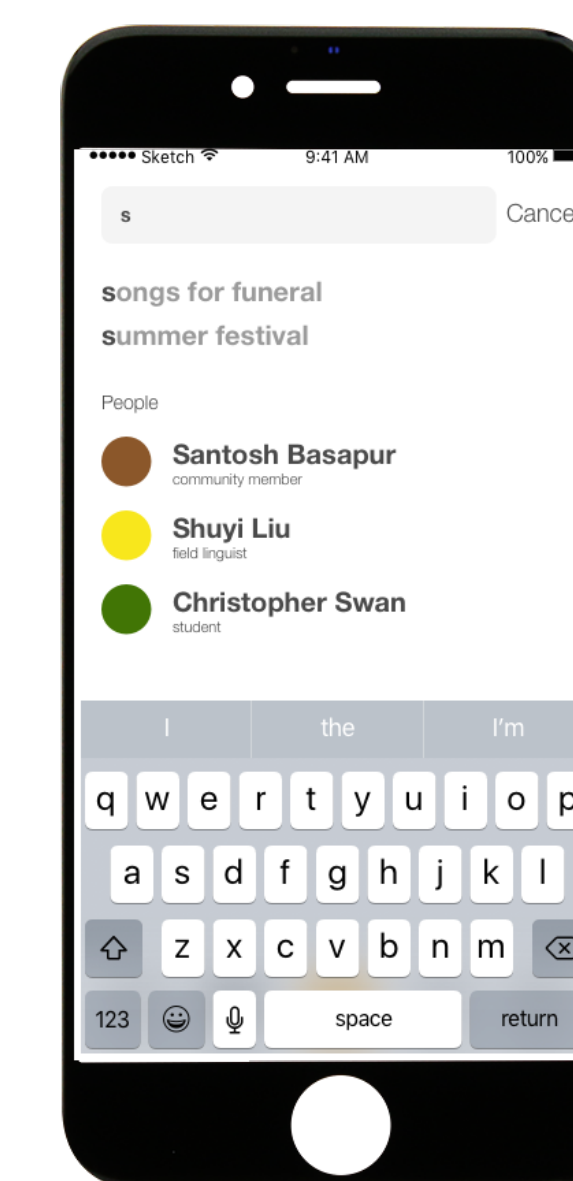
Mobile Application too supports powerful search, personal dashboard and easy data browsing for collection, annotation and deposits.



Landing Page



Personal dashboard

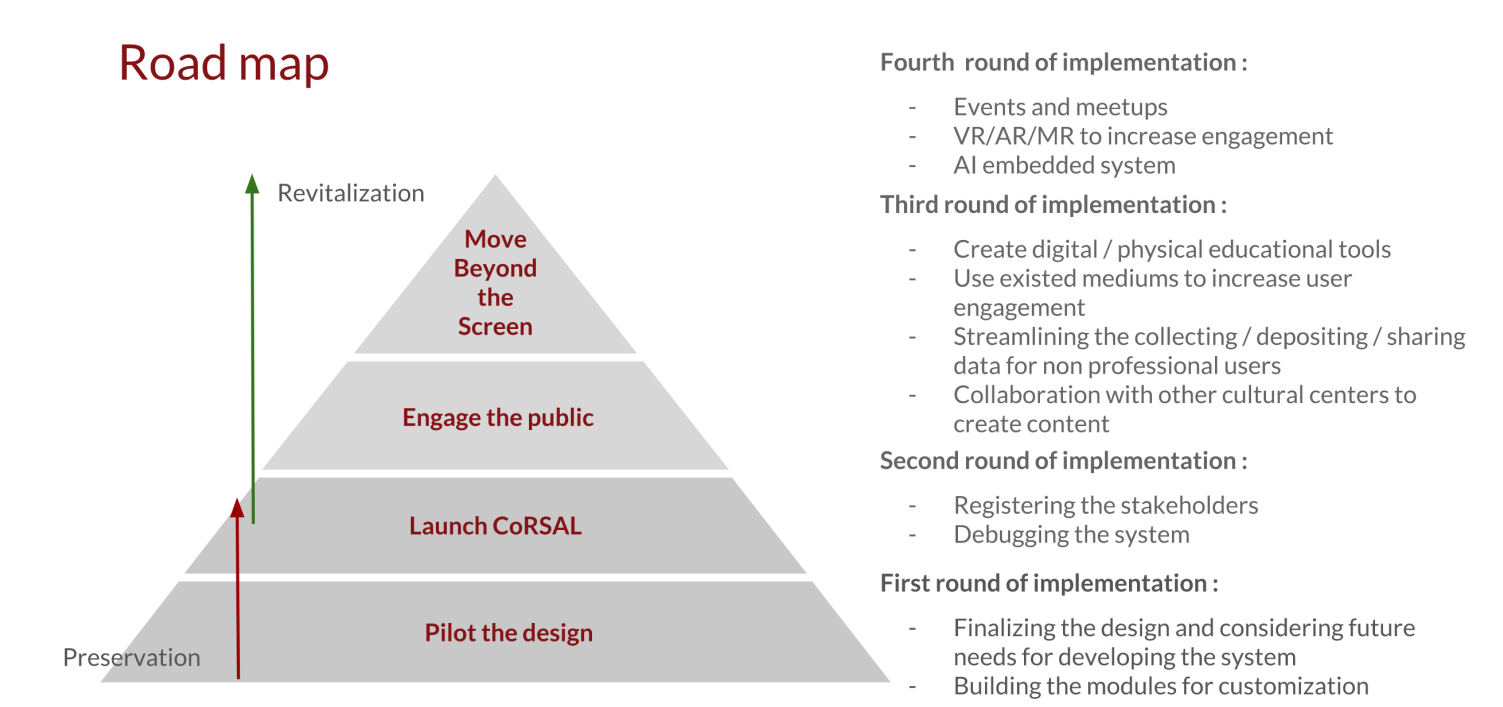
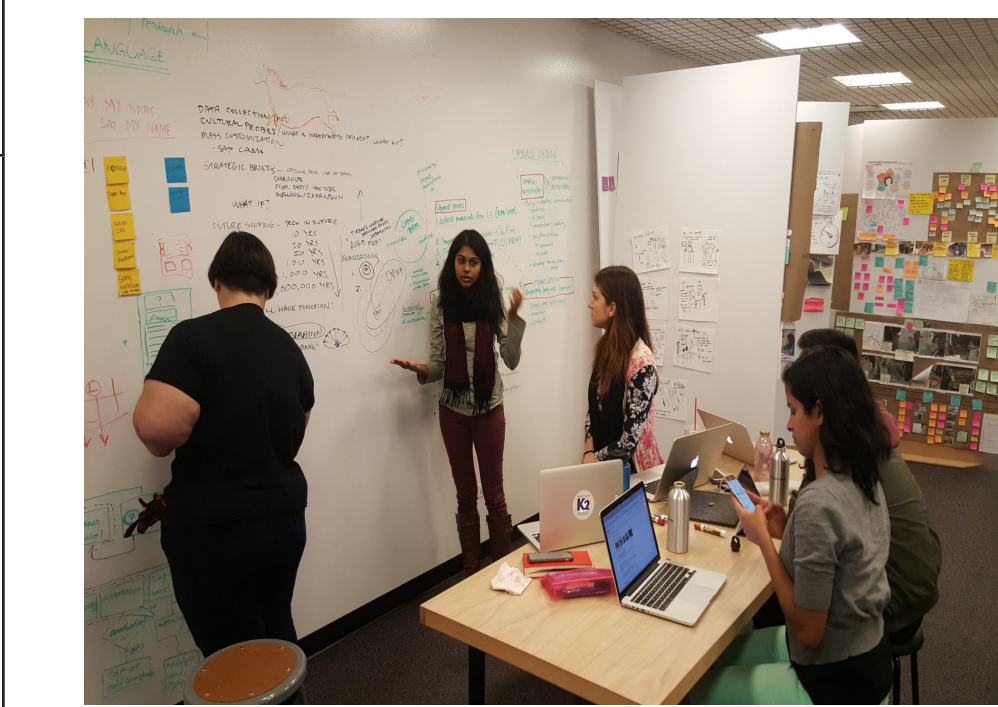


Search screen

## User Experience Design for CoRSAL

Santosh Basapur’s students at IIT’s Institute of Design created UX Design for Website and Mobile Application in Spring and Summer of 2017. User journey of potential users: Language Communities, Linguists, Computational Linguists and Depositors were mapped from the data and analysis received from UNT class. An Information Architecture, Wireframes and lastly, the look and feel of website and app was developed.

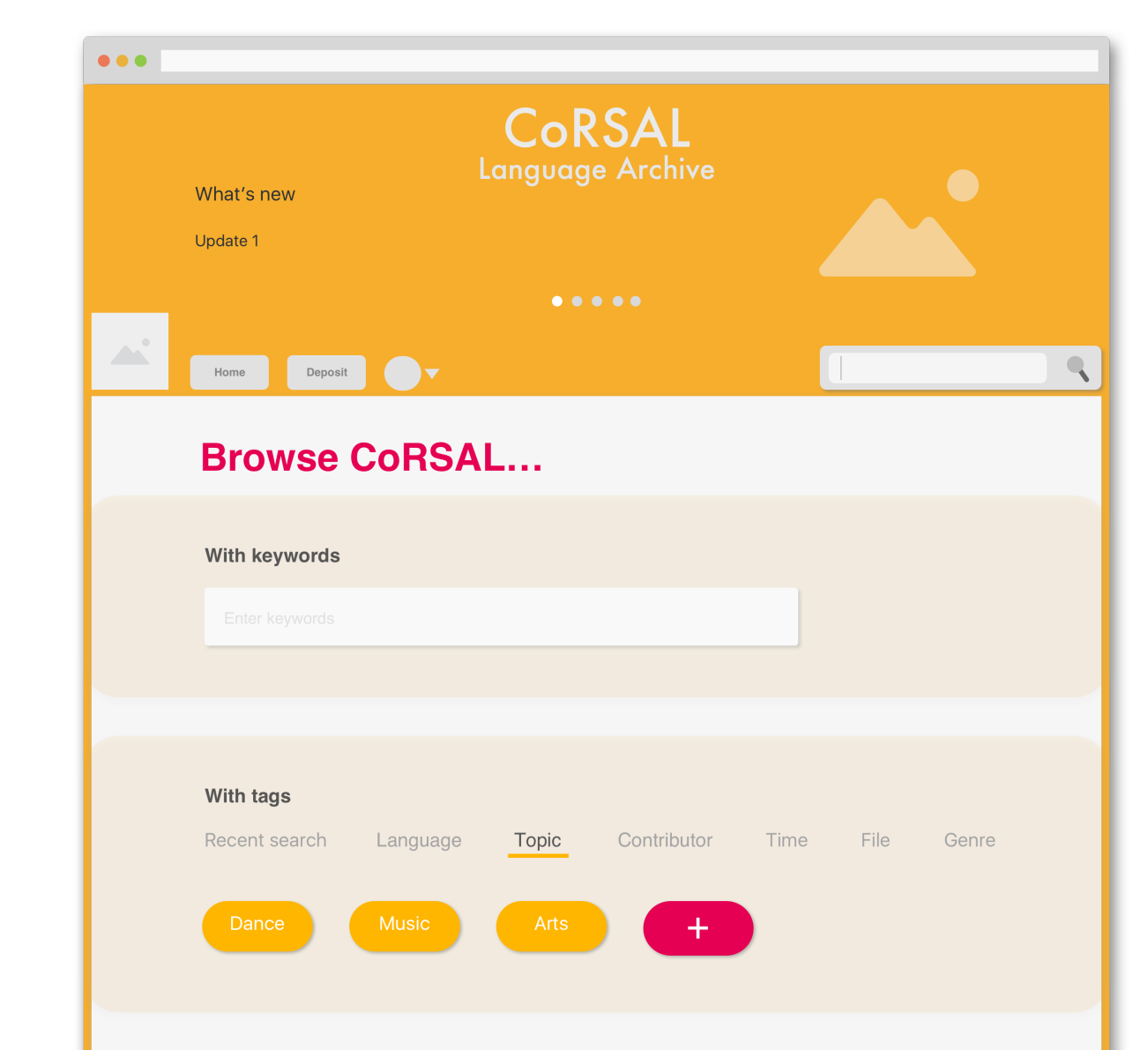
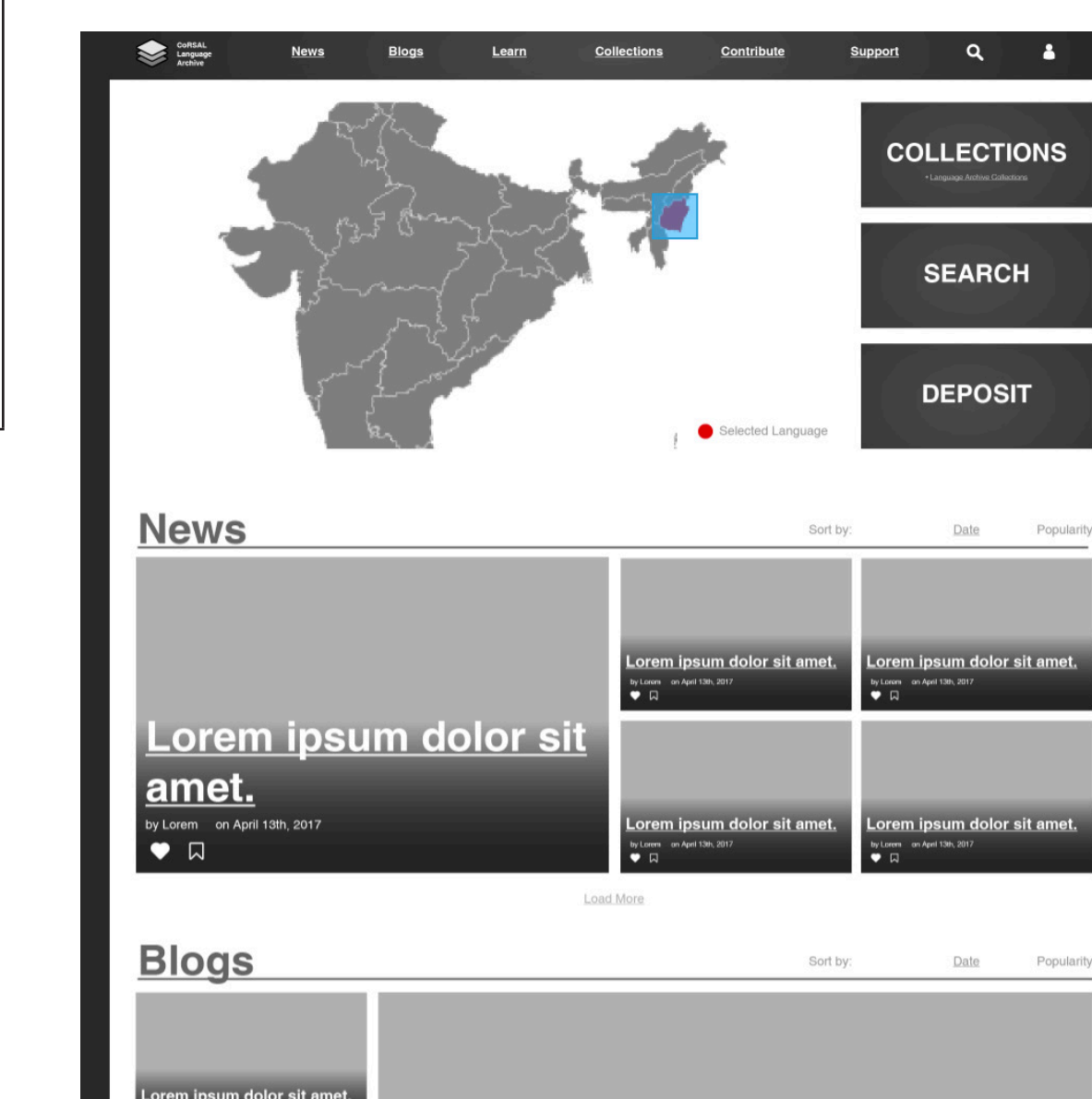
### Design Team Work Sessions



### User Journey of a Linguist



Website allows users to search, browse latest news/blog posts and multiple collections of media from depositors.



## Digital language archives of tomorrow

- Community annotations
- Community data uploads
- Crowd managed metadata
- The language archive as a “textbook” of the future
  - Diversity
  - Bilingualism
  - History
  - Language science
- Computational methods that automate annotation