

Laura Welcher, *The Rosetta Project, The Long Now Foundation*, laura@longnow.org
Fort Mason Center, Landmark Building A, San Francisco, CA 94123

The Rosetta Project Distributed Archive Model

The Rosetta Project adopts a distributed archive model for storage, discovery, and display of linguistic resources. Components of archival infrastructure reside online with third party computing services. Resources (scanned text documents, audio, video) are in the Internet Archive providing free storage, serving and robust management/metadata tools (Dublin Core, metadata harvesting protocol for Open Language Archives Community). The Internet Archive does not provide content-specific searching or browsing tools so we upload language metadata (ISO 639-3 codes, language names, taxonomic language relationships) into Google Freebase. By linking Rosetta resources in the Internet Archive to linguistic entities in Freebase we can create different tools for access and visualization.

We present a prototype wiki structured by Rosetta Freebase data, and populated by content in distributed repositories, including Rosetta's Internet Archive collection. This hybrid structured/editable wiki demonstrates the utility and power of metadata sets like OLAC, particularly linked to broader open human language datasets.