

To: [pri@ku.edu](mailto:pri@ku.edu)  
From: [stiv@u.washington.edu](mailto:stiv@u.washington.edu)  
Re: DTSL Call for papers

There are many digital resources that currently exist for language documentation, but none are suited for long term community maintained development and expansion. For example, The Linguist List is central to the discipline of Linguistics, but does not deal with language documentation directly. Digital archives and initiatives like The Rosetta Project, the Indigenous Languages of Latin America (AILLA), and Open Language Archives Community (OLAC), provide storage and growing access to digital language documentation, but do not provide tools for documenting languages. Other organizations and institutions such as the Summer Institute of Linguistics (SIL) and the Max Planck Institute (MPI) provide tools in language technology, but they maintain their own mission and development.

A major component lacking in the cyberinfrastructure of language documentation is a central repository of community maintained and updated information. In essence, we lack a Wiki of interdisciplinary knowledge and information about collaborative initiatives that share the same goal: the rapid documentation of underdocumented and endangered languages through advancements in computer technology and resource access, digital tool development, and publicity of funding opportunities. Such a repository would address a number of needs, including: interdisciplinary scholarly exchange, up-to-date information regarding tool development, advancements in hardware, digital documentation standards, practices for legacy data conversion, information about funding initiatives and study opportunities for students, pedagogical tools, and a forum for collaboration between software engineers, field linguists, and native speakers. A community maintained knowledge base could also provide information for native speaking communities that have access to the internet, as well as up-to-date information on which languages are currently being documented where, and by which researchers. There are numerous possibilities in having an extensible, accessible, and user maintained knowledge base for documenting the world's languages.

Sites like the E-MELD School of Best Practices will soon no longer be expanded due to time-constrained funding. And though initiatives like E-MELD have dramatically increased awareness of digital data obsolesce, and disseminated statements of best practice for digital documentation to the discipline, these practices will inevitably become out-of-date with technological advances. The work of E-MELD<sup>1</sup> could easily be leveraged as a starting point for a community maintained knowledge base. Migration of the site's information would include information ranging from digital documentation standards, to software and hardware recommendations for collecting and analyzing linguistic data, a large database of annotated bibliographic references, an expansive list of other web resources, and examples of language documentation projects in action. This would make a good starting point for accessible user driven content development.

By leveraging resources that have already been created, and technologies that provide users accessible and editable content, a knowledge base could be created and maintained for language documentation and digital tool development. This would provide cyberinfrastructure for collaboration among language software engineers, computational linguists, linguists, and language researchers, that is community-maintained and extensible.

---

<sup>1</sup> With permission from the E-MELD principle investigators and those involved in E-MELD.