

Digital Tools Summit for Linguistics

1. What are the most pressing needs among possible cyberinfrastructure and/or digital tools for linguistics?

Broad acceptance of existing standards (OLAC, IMDI) and extension of those standards into more finegrained levels, for example schemas for interlinear text so that the plethora of forms of interlinear text may approach a reasonable standard (like the MDF standard for fieldnames in lexical databases).

Development of tools that either do everything in one place, or else allow movement of the data in a defined workflow, with export/import options to allow lossless migration between the tools.

Establishment of properly funded repositories for linguistic data to provide incentives to researchers to prepare and deposit well-formed data.

Simple metadata entry tools based around existing standards.

Location of forms in the data via concordance linked to media to allow correction of transcripts and to assist in data analysis.

Ways of working with analysed and archived data to allow citability of both text and media.

2. What are some enduring challenges in creating cyberinfrastructure and/or digital tools for linguistics?

Lack of training for practitioners in use of tools and in understanding reasons for producing well-formed data.

Reluctance of many established linguists to engage with the possibilities offered by new technologies. Need for postgraduate programmes to take account of language documentation (which I take to include the use of appropriate digital tools and production of archival data) in presentation of dissertations.

3. Which existing resources can be leveraged to create digital tools for linguistics?

XML technologies exist generally and are being used by a small number of linguists (in the form of underlying or exportable structures in tools like Elan, Transcriber and Toolbox) and understood by an even smaller number. We have been working with Michel Jacobson at LACITO on his use of XSL to provide access to a corpus of time-aligned transcripts. The intention is to allow linguists to create files in Toolbox after having time-aligned them in an appropriate tool and then to place them into a directory from which the concordance and online viewing tool will pick them up.