

Mapping new territory: Integrating GIS into a cyberinfrastructure for linguistics

Helen Aristar Dry

Recently in the United States and Europe, there has been considerable discussion of integrated computational support for academic fields.¹ The discussion unfolds under the rubric of ‘e-science’ in Europe and ‘cyberinfrastructure’ in the United States, with the European term emphasizing the research potential of such a computational architecture, and the American term emphasizing the concept of ‘infrastructure.’ ‘Infrastructure’ has been defined as something that you use but don’t have to think about—like electricity, roads, or the telephone. However, this generation of scholars is required to think about cyberinfrastructure, because we have to build it. We have to acquire knowledge that is peripheral to our work as language analysts (e.g., knowledge of standards like XML and Unicode), so that future generations will be in the fortunate position of not having to think about the infrastructure that allows them to find, share, and analyze digital language resources.

This talk will present a non-technical overview of the interlocking web of resources, repositories, and services envisioned as the digital research environment for linguistics, and the digital standards required to implement it. It will describe the ongoing effort to develop and promulgate such standards in the linguistics community, focusing on the development of digital standards for language mapping.

The digital environment requires a great deal of “construction work,” primarily in standardization of data formats, before we can begin to pursue “normal” research within it. This is true even of what might seem conceptually simple: a map interface to language data. GIS (Geographic Information Systems) are revolutionizing some social sciences, precisely because GIS can present complex information from multiple sources in an interface so transparent that a child can understand it. No one needs instruction to understand how to click on a map to retrieve information on the languages and cultures of the area.

And yet, to produce such an interface requires the adoption, and sometimes the creation, of multiple standards new to the linguistics community. Some of these are standards which are needed for other aspects of a linguistics cyberinfrastructure, such as standard codes for language identification. Others can be adopted—although not without review—from the GIS community, e.g., standards for place identification. But others are discipline-specific: standard formats for geo-referencing language, metadata standards for language maps, and guidelines for the making of language maps themselves. Drawing on the work of the E-MELD project² and the LL-MAP project,³ this talk will discuss standards for language mapping in the context of the wider initiative to develop and promulgate the standards required by a mature cyberinfrastructure for linguistics.

¹ See, for example, recent reports on Cyberinfrastructure commissioned by the National Science Foundation (January, 2003) and the American Council of Learned Societies (December, 2006), as well as the recently instituted NSF Office of Cyberinfrastructure (<http://www.nsf.gov/dir/index.jsp?org=oci>).

² “Electronic Metastructure for Endangered Languages Data.” Funded by the National Science Foundation Social Sciences Infrastructure Division (SES-1099652), 2001-2006.

³ “Language and Location: A Map Annotation Project.” Funded by the National Science Foundation Human and Social Dynamics Division (HSD 0527512), 2005-2008.

References

- Atkins, et al. 2003. Revolutionizing Science and Engineering Through Cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure. Jan 2003 <http://www.nsf.gov/od/oci/reports/atkins.pdf>.
- Unsworth, J. et al. 2006. "Our Cultural Commonwealth: The final report of the American Council of Learned Societies Commission on Cyberinfrastructure for the Humanities & Social Sciences".