

**LREC 2014 Workshop**

**CCURL 2014**

**Collaboration and Computing for Under-Resourced Languages  
in the Linked Open Data Era**

**Reykjavík (Iceland), 26<sup>th</sup> May 2014**

**Poster Presentation**

**Title**

*Open-Source Infrastructures for Collaborative Work on Under-Resourced Languages*

**Authors**

Sjur Moshagen, Jack Rueter, Tommi Pirinen, Trond Trosterud, Francis M. Tyers

**Abstract**

In order to support crowd sourcing for a language, certain social and technical prerequisites must be met. Both the size of the community and the level of technical support available are important factors. Many language communities are too small to be able to support a crowd-sourcing approach to building language-technology resources, while others have a large enough community but require a platform that relieves the need to develop all the technical and computational-linguistic know how needed to actually run a project successfully. This article covers the languages being worked on in the Giellatekno/Divvun and Apertium infrastructures. Giellatekno is a language-technology research group, Divvun is a product development group and both work primarily on the Sámi languages. Apertium is a free/open-source project primarily working on machine translation. We use Wikipedia as an indicator to divide the set of languages that we work on into two groups: those that can support traditional crowdsourcing, and those that do not. We find that the languages being worked on in the Giellatekno/Divvun infrastructure largely fall into the latter group, while the languages in the Apertium infrastructure fall mostly into the former group. Regardless of the ability of a language community to support traditional crowdsourcing, there is in all cases the necessity to provide a technical infrastructure to back up any linguistic work. We present two infrastructures, the Giellatekno/Divvun infrastructure and the Apertium infrastructure and show that while both groups of language communities would not be able to develop language technology on their own, using the infrastructures that we present they have been quite successful.