

**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**

**Oral Presentation**

**Title**

*Crowd-Sourced, Automatic Speech-Corpora Collection – Building the Romanian Anonymous Speech Corpus*

**Authors**

Stefan Daniel Dumitrescu, Tiberiu Boroş, Radu Ion

**Abstract**

Taking the example of other successful initiatives such as VoxForge, we applied the concept of crowd-sourcing to respond to a particular need: the lack of free-speech, time-aligned, multi-user corpora for the Romanian language. Such speech corpora represent a valuable asset for spoken language processing application because they offer the means to (1) train and test acoustic models and (2) develop and validate various methods and techniques that are intended to enhance today's ASR and TTS technologies.