

**Access to data,  
The Soundbites collection of the Meertens Institute and a  
flexible approach to the curation and dissemination of  
humanities digital resources**

**Douwe Zeldenrust**

Meertens Institute (Royal Netherlands Academy of Arts and  
Sciences)

The Netherlands

douwe.zeldenrust@meertens.knaw.nl

### **Introduction**

This paper is about the curation and dissemination of humanities digital resources. The Meertens Institute, an institute of the Royal Netherlands Academy of Arts and Sciences (KNAW), studies the diversity in language and culture in the Netherlands (Meertens). It possesses a large library and numerous (audio) collections. One of these collections, The Soundbites Collection, consists of more than a thousand hours of audio recordings of Dutch dialects. The Meertens Institute started in 1950 to collect this dialect data. In the eighties collecting stopped when the recordings were sufficiently spread out over the Netherlands. Since 2009, the dialects (in total 2216 recordings) can be found on the website of the Meertens Institute (Soundbites).

### **Three interfaces**

In 2012 and 2013 additional funding was raised and substantial collections related to the original Soundbites collection have been

digitized and added. The KNAW funded a project to provide open access to the available typescripts, in total more than 11,000 scans (Archives 62, 203). Next, Data Archiving and Networked Services (DANS) funded a project to add Dutch spoken in the USA, Dutch spoken in France and the accompanying documentation (Archives Frankrijk, Daan and Marle). More than 250 hours of recordings and 2500 scans in total. The extra data created its own dynamic: to make optimal use of the possibilities of these rich resources and to cater for specific research needs, access had to be adapted. The collection is currently available via three different interfaces.

First of all, the resources have been made available through the Common Language and Resources Infrastructure, CLARIN. This opens up the resources for a wide variety of research possibilities in the Humanities and it provides dissemination, storage and the possibility of combining the resources (Zhang). The latter functionality is methodological not innovative, however, it may lead to new insights and knowledge. Using the standards of CLARIN also provides easy to use building blocks for other interfaces.

The second interface, the 'Speaking Map', is designed for phonological research of the Dutch dialects (for instance a large-scale phonetic study into vowel quality and vowel length using the combination of sound and typescript). The original website from

2009 was improved and the new interface shows the relation between the resources (Zeldenrust).

And finally a third method of disseminating, for a broad and interested audience such as secondary school students, was used. It is called the 'Language Detector' (Taaldetector). On the basis of a number of multiple-choice questions it calculates where the dialect of a speaker originates from and at the same time it offers information about the rich linguistic variation in Netherlands. The Soundbites collection is part of the Language Detector.

## **Conclusion**

The three interfaces each serve a different public: CLARIN is designed for the Humanities as a whole, the Speaking Map for phonologists and the Language Detector for a broader, general interested audience. It can be concluded that, in order to connect resources and users, the current phase of curation and dissemination of humanities digital resources asks for a flexible approach in the use of interfaces.

## References

Zeldenrust, Douwe and Marc van Oostendorp (2013), Combining tailor made research solutions with big infrastructures: The speaking map of the Netherlands, in: *DH2013 conference abstracts* (University of Nebraska).

Zhang, Junte Marc Kemps-Snijders, and Hans Bennis (2012). The CMDI MI Search Engine: Access to Language Resources and Tools Using Heterogeneous Metadata Schemas, in: *Lecture Notes in Computer Science, 2012, Volume 7489, Theory and Practice of Digital Libraries, Pages 492-495.*

### *Archives*

The Archives of the Meertens Institute, collection no. 62.

The Archives of the Meertens Institute, collection no. 203.

The Archives of the Meertens Institute, Nederlands in Frankrijk, (1964 - 1966).

The Archives of the Meertens Institute, Nederlands in de USA, collectie Jo Daan, (1966).

The Archives of the Meertens Institute, Nederlands in de USA, collectie van Marle, (1994 -1998).

### *Websites*

[www.clarin.eu](http://www.clarin.eu) (Accessed January 31, 2014).

[www.meertens.knaw.nl](http://www.meertens.knaw.nl) (Accessed January 31, 2014).

[www.meertens.knaw.nl/cms/en/component/content/article/3-nieuws/143948-meertens-instituut-lanceert-taaldetector](http://www.meertens.knaw.nl/cms/en/component/content/article/3-nieuws/143948-meertens-instituut-lanceert-taaldetector) (Accessed January 31, 2014).

[www.meertens.knaw.nl/soundbites](http://www.meertens.knaw.nl/soundbites) (Accessed January 31, 2014).