Alexander Geyken and Thomas Gloning

A living text archive of 15th–19th c. German. Corpus strategies, technology, organization

The corpus situation for the study of Early New High German (ENHG) and the older stages of New High German (oNHG) is far from satisfying. Indeed, there is no integrated and balanced corpus of a sufficient size that is publicly available. In this paper we give an outline of a living text archive of ENHG and of oNHG in a distributed infrastructure that is supposed to fill this gap in the coming years. Such an archive is designed as a collaborative, sustainable and interoperable platform where historical texts are integrated together with relevant metadata and expert computational technology. We briefly mention examples of research questions and of corpus building scenarios and comment on the organizational aspects of such an archive including its potential as a basis for reference corpora for specific purposes.