Abstract. Latvian National Corpora Collection (LNCC) is a diverse collection of corpora representing both written and spoken language. LNCC covers various use cases and all the important text types and genres. It is a continuous multi-institutional and multi-project effort, supported by the Digital Humanities and Language Technology communities in Latvia. All corpora of LNCC are annotated with a uniform morpho-syntactic annotation scheme which enables federated search and consistent linguistics analysis in all the LNCC corpora and allows to select and mix various corpora for pre-training large Latvian language models like BERT and GPT.