CEDAR: The Dutch Historical Censuses as Linked Open Data

Albert Meroño-Peñuela albert.merono@vu.nl
VU University Amsterdam & Data Archiving and Networked Services

Ashkan Ashkpour ashkan.ashkpour@iisg.nl
International Institute of Social History

Christophe Guéret c.d.m.gueret@vu.nl
VU University Amsterdam & Data Archiving and Networked Services

Stefan Schlobach k.s.schlobach@vu.nl
VU University Amsterdam

Abstract

In this document we describe the CEDAR dataset, a five-star Linked Open Data representation of the Dutch historical censuses, conducted in the Netherlands once every 10 years from 1795 to 1971. We produce a linked dataset from a digitized sample of 2,288 tables. The dataset contains more than 6.8 million statistical observations about the demography, labour and housing of the Dutch society in the 18th, 19th and 20th centuries. The dataset is modeled using the RDF Data Cube vocabulary for multidimensional data, uses Open Annotation to express rules of data harmonization, and keeps track of the provenance of every single data point and its transformations using PROV. We link these observations to well known standard classification systems in social history, such as the Historical International Standard Classification of Occupations (HISCO) and the Amsterdamse Code (AC), which in turn link to DBpedia and GeoNames. The two main contributions of the dataset are the improvement of data integration and access for historical research, and the emergence of new historical data hubs, like classifications of historical religions and historical house types, in the Linked Open Data cloud.