



Royal Netherlands Academy of Arts and Sciences (KNAW) KONINKLIJKE NEDERLANDSE AKADEMIE VAN WETENSCHAPPEN

Integrating Historical Census Data in the Semantic Web

Meroño-Peñuela, Albert; Ashkpour, Ashkan

2014

document version

Early version, also known as pre-print

[Link to publication in KNAW Research Portal](#)

citation for published version (APA)

Meroño-Peñuela, A., & Ashkpour, A. (2014). *Integrating Historical Census Data in the Semantic Web*.
<http://historicalnetworkresearch.org/hnr-conference-2014/>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the KNAW public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain.
- You may freely distribute the URL identifying the publication in the KNAW public portal.

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:

pure@knaw.nl

Integrating Historical Census Data in the Semantic Web

Albert Meroño-Peñuela^{1,2} and Ashkan Ashkpour^{3,4}

¹ Department of Computer Science, VU University Amsterdam, NL

² Data Archiving and Networked Services, KNAW, NL
albert.merono@vu.nl

³ Erasmus University Rotterdam, NL

⁴ International Institute of Social History, NL
ashkan.ashkpour@iisg.nl

Abstract. Historical censuses are one the most consulted, reliable and large scale statistical data sources available, describing the demographic, social and economic history of a nation. To answer their research questions, researchers often need to query long time series of census data. However, such longitudinal queries are typically hampered by the scarce integration of the historical censuses, demanding manual and knowledge intensive harmonization and restructuring in order to obtain meaningful comparisons over time. The challenges are even harder if provenance microdata is lost. In this paper we describe the methodology followed in CEDAR⁵, a project of the Computational Humanities Programme⁶, to provide solutions to these data-issues in the Dutch historical censuses (1795-1971). Our proposal builds on top of Linked Data and the Resource Description Framework (RDF) technology, allowing us to transform the original census tables into a graph of Linked Census Data. With such a graph, every census data-item can be interlinked on the Web with other hubs of historical socioeconomic and demographic information. By following the Linked Data principles, our aim is two-fold. On the one hand, we show how the integration of our own historical census data is improved by linking them to the network of Linked Historical Datasets on the Web. On the other hand, we envisage new historical classifications (like demographical structures, housing types, occupational classes and statuses, or religious denominations) coming out of our harmonization process, which are not published yet on the Web on a standard manner and could improve the interoperability of other datasets.

⁵See <http://www.cedar-project.nl/>

⁶See <http://www.ehumanities.nl/>