BRILL

RESEARCH DATA JOURNAL FOR THE HUMANITIES AND SOCIAL SCIENCES (2018) 1-14

BRILL OPEN
brill.com/rdj

Improving Access to the Dutch Historical Censuses with Linked Open Data

Social and Economic History

Albert Meroño-Peñuela
Vrije Universiteit Amsterdam, Netherlands
albert.merono@vu.nl

Ashkan Ashkpour
International Institute of Social History (IISH), Amsterdam, Netherlands
ashkan.ashkpour@iisg.nl

Valentijn Gilissen
Jan Jonker
Tom Vreugdenhil
Peter Doorn
Data Archiving and Networked Services (DANS), Royal Netherlands Academy of Arts and Sciences (KNAW), The Hague, Netherlands
valentijn.gilissen@dans.knaw.nl
jan.jonker@dans.knaw.nl
tom.vreugdenhil@dans.knaw.nl
peter.doorn@dans.knaw.nl

Abstract

The Dutch Historical Censuses (1795–1971) contain statistics that describe almost two centuries of History in the Netherlands. These censuses were conducted once every 10 years (with some exceptions) from 1795 to 1971. Researchers have used its wealth of demographic, occupational, and housing information to answer fundamental questions in social economic history. However, accessing these data has traditionally been a time consuming and knowledge intensive task. In this paper, we describe the outcomes of the CEDAR project, which make access to the digitized assets of the Dutch Historical Censuses easier, faster, and more reliable. This is achieved by using the data publishing paradigm of Linked Data from the Semantic Web. We use a digitized sample of 2,288

census tables to produce a linked dataset of more than 6.8 million statistical observations. The dataset is modeled using the RDF Data Cube, Open Annotation, and PROV vocabularies. The contributions of representing this dataset as Linked Data are: (1) a uniform database interface for efficient querying of census data; (2) a standardized and reproducible data harmonization workflow; and (3) an augmentation of the dataset through richer connections to related resources on the Web.

Keywords

census data – historical statistics – semantic web – linked data – data harmonization

– Related data set "Dutch Historical Censuses" with URL www.doi.org/10.17026/dans-xpk-wj5w in repository "DANS".

1. Introduction

The Dutch historical censuses were conducted 17 times from 1795 until 1971, once every 10 years. For each of these, the government counted the entire population of the Netherlands, and aggregated the results in three different censuses: demographic (with variables such as gender or age), occupational (about jobs of citizens), and housing (about public and private buildings and other living facilities). After 1971, this exhaustive data collection stopped mostly due to social opposition, and authorities switched to municipal registers and sampling. Nevertheless, the data collected in the 1795–1971 period is of special interest to historians and social scientists because of three facts: (1) it is based on counting the whole Dutch population, instead of sampling; (2) it provides an unprecedented level of detail, hardly comparable to modern censuses due to privacy regulations; and (3) the survey microdata from which the aggregations were originally built is almost entirely lost.

The 1795–1971 census results were published in books containing the aggregated statistical tables, and several institutes, among which the Central Bureau of Statistics[1] (CBS) and the International Institute of Social History[2] (IISH), are holding paper copies. Offering access to these data in a systematic way has always been a priority in these and other institutions. In an effort to improve this access, part of the tables in the historical censuses books have been

1 See https://www.cbs.nl/.

2 See https://socialhistory.org/.

digitized as 300,000 scanned images[3] in various projects between the CBS, the IISH and several institutes of the Royal Netherlands Academy of Arts and Sciences[4] (KNAW), such as Data Archiving and Networked Services[5] (DANS) and the Netherlands Interdisciplinary Demographic Institute[6] (NIDI). In addition, these projects have translated part of these scans, by manual input, into more structured formats, resulting in a collection of 507 machine-readable Excel spreadsheets, containing 2,288 census tables.[7]

In this paper, we describe the digitally archived results of converting these machine-readable census spreadsheets into Linked Data, a paradigm for publishing structured data on the Web (Heath & Bizer, 2011), to which we refer as the CEDAR RDF Database. The CEDAR RDF Database contains the final database of the harmonized Dutch historical censuses, encoded using the Resource Description Framework (RDF) in two variants: as a complete RDF conversion of the 2,288 tables, although partially harmonized, of the 1795–1971 period (the "cedar" variant); and a partial RDF conversion of 140 tables, fully harmonized, of the 1859–1920 period (the "cedar-mini" variant). This database, its variants, and their associated documentation and metadata are available online in many forms, including a SPARQL endpoint;[8] deposited at the DANS archiving system EASY[9] and in the form of a website.[10]

Rather than pursuing the answer of a specific research question, the construction of the CEDAR RDF Database sets the foundation of a platform for researchers and practitioners to facilitate discovering and answering their own research questions. This platform consists of the dataset itself and a set of services on top of it. These services enable the access to Dutch historical registers for users in an unprecedented longitudinal and comparable way, overcoming to a large extent the historical difficulties of analyzing census data across time. In addition, these services use the archived census source materials to provide accurate provenance information; and minimize the efforts required from users in data cleaning and data preparation.

3 The number of statistical tables contained in these scans, as well as the total number of tables contained in the printed books, have never been counted and hence remains unknown.

4 See http://knaw.nl/.

5 See https://dans.knaw.nl/.

6 See http://nidi.knaw.nl/.

7 See http://volkstellingen.nl/.

8 See http://lod.cedar-project.nl/cedar/sparql and http://lod.cedar-project.nl/cedar-mini/sparql.

9 See https://easy.dans.knaw.nl/ui/datasets/id/easy-dataset:65225.

10 See http://www.censusdata.nl.

2. Problem

The Dutch historical censuses were collected with different information needs at given times, implying deep changes in their structure, codes, variables and survey questions (Ashkpour, Meroño-Peñuela & Mandemakers, 2015). However, these changes make comparisons across time hard, and consequently the use of the historical censuses for longitudinal analysis cumbersome. Addressing this requires extensive manual input from a domain expert. Moreover, this data munging is repeated over and over by different researchers when they start new research. Concretely, previous research (Ashkpour, Meroño-Peñuela & Mandemakers, 2015) shows that the access problems of this dataset are:

- **Aggregated data.** The original surveys describing individuals are not available, and only the aggregated census results remained. This hampers a longitudinal harmonization of census data across multiple years, since variables aggregated differently are hardly comparable. For example, *the number of female bakers in Haarlem in 1849 and 1859* might be incomparable if the census of 1849 counted bakers without distinguishing gender, and the census of 1859 only counted female bakers in Noord-Holland (the province where Haarlem is).
- **Changing variables.** The possible values that a variable can get vary over time. Although some census variables are completely stable (for example, the possible values of *sex* are always *male*, *female* and *unknown*), some others are very dynamic (for example, the many possible, epoch-dependent values of *occupation*).
- **Missing codes.** As a consequence of the previous point, a variable in a specific census year might have an aggregation level that does not have an equivalent in another census year. For example, for the classification of housing type, we have very specific codes for counting people in barracks (e.g., *Kazerne der Marechaussee*, *Artilleriekazernes* and so on) or forts (e.g., *Fort Isabelle* and *Fort Kijkduin*); as we do not have this detailed information for all years, we need to aggregate these housing types according to their function into the higher code *Military Buildings*.
- **Structural heterogeneity.** The collection is curated to be visually faithful with respect to the book originals. This means that the spreadsheet layouts are historically coherent, but also very difficult to query in an homogeneous way.
- **Inconsistency.** Data quality is an important issue within this dataset, and data errors have been previously detected (Ashkpour, Meroño-Peñuela & Mandemakers, 2015). We distinguish between two different kinds of data

errors: those already present in the source books, and those introduced by the digitization process. Typically these are in the form of spelling mistakes and variants, contents of columns which have shifted to another column, and columns wrongly merged during transcription. Numeric errors are usually detected by statistical methods, e.g. by checking conformance to Benford's Law (Benford, 1938).

– **Non reproducibility.** Perhaps more importantly, and as a consequence of its poor and loosely connected data representation paradigms (either books, scanned images, or digital spreadsheets), previous socio-historical research that uses this dataset as a source is hardly reproducible. To improve this reproducibility, new paradigms for representing both the original data, and the processes that affect those data, at a very fine-grained level are needed.

3. Methods

Figure 1 shows the integration pipeline that we used to convert the Dutch historical censuses dataset into the CEDAR RDF Database, a fully fledged, 5-star Linked Dataset. This process starts at the left of Figure 1, where the original spreadsheets are retrieved from their authoritative archive.

The first step after retrieving the original spreadsheets is *cell markup*. In it, a group of trained experts need to annotate these spreadsheets with a style/color code, producing tables like the one shown in Figure 2. This style/color markup is used to distinguish key information provided in the table, identifying *row properties* (i.e. variable names, like *sex* and *age*), *column* and *row headers* (i.e. variable values, like *male* and *female*), and *data* (such as *8 people*). This markup is used in the next step to correctly interpret the observations contained in the table.

The second step uses a program called TabLinker[11] to read the spreadsheets and their markup, and produce what we call the *raw data* representation of the dataset. This raw data is already a graph database in RDF format. To produce it, TabLinker uses the original spreadsheets and their markup to interpret *observations*: a specific population count that depends on a number of variable names and values. For instance, the bottom right figure in Figure 2 is interpreted as the following observation: there were 12 unmarried (*O* column) women (*V* column), 12 years old and born in 1878 (*12 1878* column) working as ordinary (row *D* in column *Positie in het beroep*, position in the occupation)

11 See https://github.com/Data2semantics/TabLinker.

diamond cutters (*Diamantsnijders* row) in the municipality of Amsterdam (column *Gemeente*, municipality). However, this raw data representation is still not homogeneously queryable due to the lack of harmonization of variable names and values.

In the third step, we combine systems and expert knowledge to harmonize the variables, values and classification systems of the census, using a set of mappings. Our harmonization approach builds on a flexible, yet structured, workflow. This workflow allows researchers to iteratively discover the peculiarities of such a challenging dataset and provide (different) interpretations on the data in an accountable way (Ashkpour, Mandemakers & Boonstra 2016). In order to do so we propose an approach called source-oriented harmonization. By making the harmonization process more structured and explicit we aim stimulate similar efforts across other datasets and projects. The produced mappings (outcome of the harmonization) are *rules* that establish how a specific string in the original tables (e.g. *Diamantsnijders*) should be represented as a unique, standard, harmonized code among the whole dataset (e.g. *hisco:88030*).[12] The mappings also indicate the variable to which they must be applied (e.g. *occupation*). Multiple classification systems are used in these mappings, depending on their level of genericity and availability. To reuse as many as we can, and also to assess how globally in use they are, we use

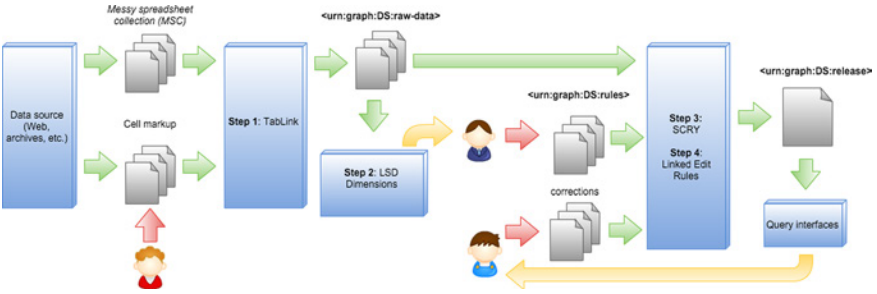

FIGURE 1 *Integration pipeline for the CEDAR RDF Database. The workflow starts at the archiving system, where the original Excel files are stored and retrieved using its API. Raw data is produced after interpreting complex table layout. These raw data are later transformed into harmonized data by applying integration rules encoded as Open Annotations. Red arrows indicate that manual input is required.*

12 See http://historyofwork.iisg.nl for a description of HISCO, the Historical International Standard Classification of Occupations (van Leeuwen, Maas & Miles, 2002).

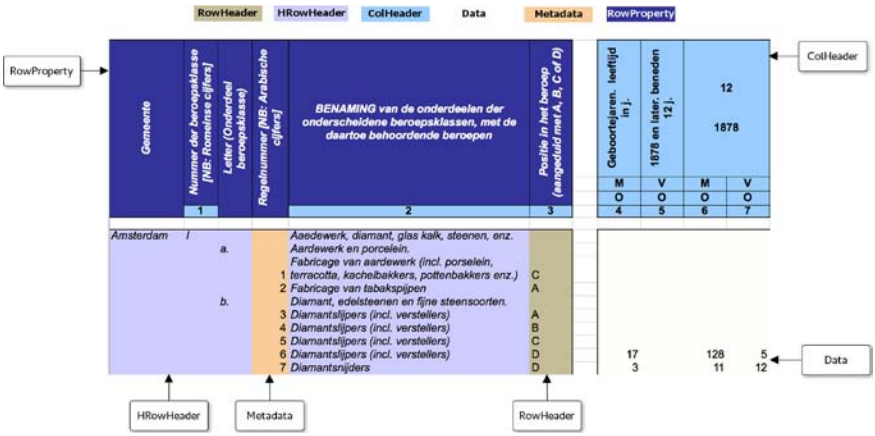

FIGURE 2 *One of the census tables of the dataset (occupation census of 1889, province of Noord-Holland). Colour markup is manually added and does not belong to the original data.*

LSD Dimensions[13] (Meroño-Peñuela, 2014). LSD (for Linked Statistical Data) Dimensions is an index of statistical variables and values currently published as Linked Data on the Web, where users can find standard identifiers for variables, such as "sdmx:sex", and their corresponding values, such as "sdmx:sex-male" and "sdmx:sex-female". This way, we reuse many variable names and code lists such as gender, age, reference area (place), reference period (time), and so on, mostly based on their SDMX[14] (Statistical Data and Metadata eXchange) counterparts.

In the last step, the execution of these *rules* over the *raw data* yields what we call the *release* data, a 4-star Linked Dataset that can be homogeneously queried, and whose results can be consistently replicated.

4. Data

- **Dutch Historical Censuses deposited at DANS** – DOI:10.17026/dans-xpk-wj5w
- **Temporal coverage:** 1795–1971

13 See http://lsd-dimensions.org/.

14 See https://sdmx.org/.

The output of the presented method generates 6,800,175 census observations. The distribution of variables (dimensions) among these observations is shown in Table 1.

The CEDAR RDF Database is released in two different variants: "cedar" and "cedar-mini". The "cedar" variant contains an RDF conversion of the complete 1795–1971 data series, although only partially harmonized. The "cedar-mini" variant contains an RDF conversion of the shorter period 1859–1920, but with highly curated harmonization mappings. For a complete description of the differences between the two subsets, we refer the readers to (Meroño-Peñuela, Ashkpour, Guéret & Schlobach, 2015) and to the dataset documentation in EASY.

The dataset is modeled using the RDF Data Cube,[15] Open Annotation,[16] and PROV[17] vocabularies. These schemas are specifically designed to publish statistical data, annotation data, and provenance data on the Web as Linked Data, respectively. We choose these vocabularies due to their fit with the requirements of the CEDAR RDF Database. RDF Data Cube offers an appropriate representation of statistical observations that matches the interpretation of the Dutch census tables by experts (see Figure 2). We use Open Annotations to

TABLE 1 *Dimensions of the dataset; the second column indicates how many observations in the dataset refer to such dimension, the third column indicates the proportion of observations referring to such dimension with respect to the total number of observations (6.8M)*

Dimension	**Occurrences**	**% obs.**
Belief	253,480	3.73%
Census type	4,642,360	68.27%
Municipality	153,248	2.25%
Marital status	1,886,415	27.74%
Occupation	328,790	4.84%
Occupation position	8,120	0.12%
Province	43,946	0.65%
Reference period	4,642,360	68.27%
Sex	3,801,431	55.90%

15 See https://www.w3.org/TR/vocab-data-cube/.

16 See https://www.w3.org/TR/annotation-vocab/.

17 See https://www.w3.org/TR/prov-primer/.

make statements about *missing codes*, table layouts (*structural heterogeneity*), and data *inconsistencies* (see Section "Problem"), without modifying the original contents of the tables. Finally, we use PROV to describe data provenance and all data transformations we perform, from the original table values to their final harmonized form. This allows us to address the issues on *aggregated levels*, *changing variables*, and *reproducibility* (see Section "Problem"), since users can not only make use of the data, but also follow the trails of how it came to be. The reproducibility of analyses on this dataset is further enhanced by the availability of publicly accessible tables that can be linked and referenced, rather than having researchers copying and typing their own data for each research study.

Remarkably, the CEDAR RDF Database can be queried in a straightforward way for data that can be used to support an hypothesis or an answer to a meaningful historical question (see Section "Data usability & New possibilities"). After executing the relevant queries, these data are typically provided in the order of seconds; the process of gathering them before the existence of the CEDAR RDF Database implied manual labour in the order of hours or even days. Table 2 shows examples of such queries, the number of tables implied in resolving them, and the number of cells used to construct the results. These queries, and several additional ones, can be run online at http://lod.cedar-project.nl/cedar/data.html.

According to the Design Principles of Linked Data,[18] a Linked Dataset is only "5-star Linked Data" if its resources are connected to external Linked Datasets as well. In order to to make the CEDAR RDF Database a 5-star Linked Dataset, we issue links as depicted by Figure 3. Concretely, we use expert knowledge to link the census data to: (1) occupations in the the Historical International Standard Classification of Occupations (HISCO) (van Leeuwen, Maas & Miles, 2002); (2) occupations in the ICONCLASS[19] system, with alternative occupation descriptions; (3) to historical Dutch municipalities in gemeentegeschiedenis.nl (Zandhuis, den Engelse & Mac Gillavry, 2015), which link to DBpedia, GeoNames, codes of the CBS (Centraal Bureau voor de Statistiek), and codes of the Amsterdamse Code (van der Meer & Boonstra, 2006); and to maritime trade registers of the Dutch Ships and Sailors dataset (de Boer, Leinenga, van Rossum & Hoekstra, 2014).

Additional technical documentation to deploy, reproduce, query, refer and extend the CEDAR RDF Database can be found at the EASY deposit metadata,[20]

18 See https://www.w3.org/DesignIssues/LinkedData.html.

19 See http://iconclass.org.

20 See https://easy.dans.knaw.nl/ui/datasets/id/easy-dataset:65225.

TABLE 2 *Example queries that can be answered using the cedar-mini subset, with the number of tables and cells used to answer them*

Query	#Tables	#Cells
Inhabited houses in Zuid-Scharwoude in 1889	1	1
Occupied houses and living ships per municipality	59	80,032
Legally registered and present inhabitants per municipality	34	23,086
Houses under construction	47	4,478
Empty houses	59	34,834
Temporarily present inhabitants in ships	35	4,255
Temporarily present inhabitants per municipality	47	74,462
Temporarily absent inhabitants per municipality	34	37,044
Temporarily present inhabitants in wagons	13	426
Number of houses according to their type, from 1859 until 1920	59	136,768

in the GitHub CEDAR-project organization,[21] in the technical papers cited in the references section, and at http://lod.cedar-project.nl/.

4.1. *Data Usability & New Possibilities*

To improve the usability of the resulting dataset, we make available various interfaces for accessing it. The simplest of these is a website which makes all data available for download.[22] For more fine-grained data querying needs, we have set a SPARQL endpoint where all the data can be queried using the SPARQL query language.[23] We also make publicly available a number of example SPARQL queries.[24] However, use these queries requires Linked Data expertise, and knowledge of the SPARQL query language. To address this, we also publish a Linked Data API and a query frontend using recent tools (Meroño-Peñuela & Hoekstra, 2017).[25] This interface allows both humans and machines to access the data in a systematic way. For instance, the query *houseType_all* allows to query all house-related information, while *houseType_params* allows to specify concrete values for the various variables in the query.

21 See https://github.com/CEDAR-project/.

22 See http://censusdata.nl/.

23 See http://lod.cedar-project.nl/cedar/sparql and http://lod.cedar-project.nl/cedar-mini/sparql.

24 See https://github.com/CEDAR-project/Queries.

25 See http://grlc.io/api/CEDAR-project/Queries.

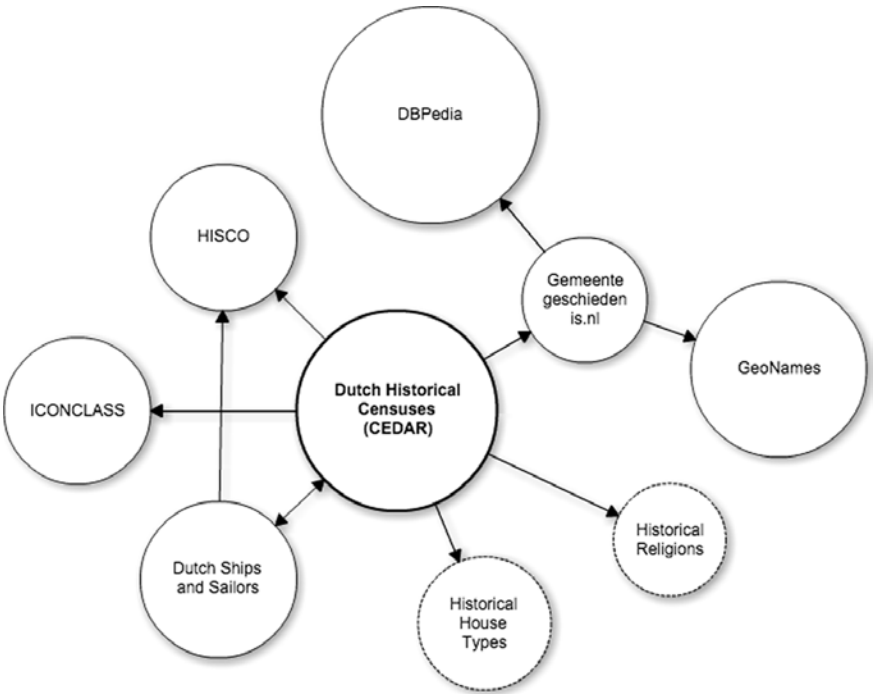

FIGURE 3 *External datasets in the Linked Open Data (LOD) cloud to which the CEDAR RDF Database is connected to, achieving the 5th Linked Data star.*

The resulting database, and its outgoing links, open up for old and new research possibilities, some of them already in place. For instance, NLGIS,[26] a web portal to display Dutch demographics of the 19th century in maps, makes use of our data to extend their coverage to variables only present in the historical censuses. But perhaps the most immediate potential of the CEDAR database is the ability to reproduce the results of existing historical research, and use queries to build the same data that researchers have used in the past to answer their research questions. For example, (Boonstra, Doorn, van Horik, van Maarseveen & Oudhof, 2007) offers an invaluable collection of data-driven research on the digital version of the censuses, each of its chapters addressing one specific historical research question. The data used in these chapters could now be easily reconstructed by using the CEDAR database and its provided queries and APIs. For example, in the chapter *"Beter wonen? Woningmarkt en residentiële segregatie in Amsterdam 1850–1940"*, H.M. Laloli investigates whether, and the extent to which, the quality of housing in Amsterdam increased in the period 1850–1940. To do so, he gathers data of the Dutch historical censuses

26 See http://nlgis.nl/.

of this period about houses, in particular on *Bewoonde* and *Onbewoonde huizen* (occupied or unoccupied houses, Table 1, pp. 160). These variables are included in the CEDAR database, and the original cross-year harmonized tables can be rebuilt through its API.[27] Another example research question in (Boonstra, Doorn, van Horik, van Maarseveen & Oudhof, 2007) looks at what degree did economic specialization occur in the Netherlands in the 19th and 20th centuries. Here, the mapping with HISCO is a fundamental but time-consuming task that now can be done and further refined with a more direct access to the data.[28] This ease at reproducing existing studies can also be understood as an opportunity for extending them and augment their coverage. For example, census data on the historical Dutch women labour force have been used in comparative studies with other nations, finding that Dutch women might have had a higher participation in the labour market than previously suggested (Schmidt & van Nederveen Meerkerk, 2012). Although it would require for other sources to use Linked Data as data representation paradigm, the extension of this study to cover databases of other countries would be not only feasible, but only require to use the same queries over different datasets; a practice that we see already happening in other areas of socioeconomic history (Hoekstra et al., 2018).

In addition to reproducing old research questions, new research questions suggested by semantically rich links can now be more easily investigated: for instance, what is the relationship between exported goods in ships and their manufacturing industries in the Netherlands during the 18th and 19th centuries? Or: to what extent did the demographic characteristics of artists and artisans change after the Golden Age in different disciplines? One of the most interesting methodological outcomes is the *transposability* of research questions: the same queries can be reused and executed over different sets of the data by just changing one parameter, which can be useful in comparative studies.

5. Concluding Remarks

The Dutch historical census dataset is surrounded by a history of its own, where many have devoted life-long efforts in improving the access to the most important collection of historical statistics about the past of the Netherlands. This paper summarizes the CEDAR RDF Database, an archived dataset that

27 See e.g. the query http://grlc.io/api/CEDAR-project/Queries/houseType_all.

28 See the HISCO query API at http://grlc.io/api/CLARIAH/wp4-queries-hisco/.

solves some issues related to the accessibility of these data, in particular dealing with their harmonization, querying and reproducibility of studies. This only adds one step to a long data curation tradition, and the authors hope that it will be only the first of many still to come.

Many challenges are open for the future, but we have a special interest in investigating the genericity of our methods, as well as their applicability to other datasets. As mentioned in our contributions, we will transpose the same research questions in form of Linked Data queries to other linked datasets, to investigate if these queries can be re-executed on different data with minimal changes. With respect to the applicability of our methodology to other datasets, we have already successfully published a number of statistical tables as Linked Data by following the same workflow;[29] but in order to strengthen this point within the same domain we will aim at other international census publications, such as the New Zealand censuses[30] and Britain's Histpop reports.[31]

References

All hyperlinks accessed on 2018-06-15.

Ashkpour, A., Meroño-Peñuela, A., Mandemakers, K. The Aggregate Dutch Historical Censuses: Harmonization and RDF. Historical Methods: A Journal of Quantitative and Interdisciplinary History, 48(4), pp. 230–245, 2015.

Ashkpour, A., Mandemakers, K., Boonstra, O. Source Oriented Harmonization of Aggregate Historical Census Data: A Flexible and Accountable Approach in RDF. Historical Social Research, 41(4), pp. 291–321, 2016.

Benford, F. The law of anomalous numbers. Proceedings of the American Philosophical Society 78 (4), pp. 551–72, 1938.

de Boer, V., Leinenga, J., van Rossum, M., Hoekstra, R. Dutch Ships and Sailors Linked Data Cloud. In: The Semantic Web – ISWC 2014. 13th International Semantic Web Conference, Riva del Garda, Italy, October 19–23, 2014, LNCS 8796, pp. 229–244. Springer, 2014.

29 See e.g. http://albertmeronyo.github.io/prices-wages/ and http://albertmeronyo.github.io/uk-messy/.

30 See http://m.stats.govt.nz/browse_for_stats/snapshots-of-nz/digitised-collections/census-collection.aspx.

31 See http://www.histpop.org.

Boonstra, O., Doorn, P., van Horik, M., van Maarseveen, J., Oudhof, J. Twee eeuwen Nederland geteld. Onderzoek met de digitale Volks-, Beroeps- en Woningtellingen 1795–2001. DANS – Data Archiving and Networked Services en CBS, The Hague, 2007.

Heath, T. , Bizer, C. Linked Data: Evolving the Web into a Global Data Space (1st edition). Synthesis Lectures on the Semantic Web: Theory and Technology, 1(1), pp. 1-136. Morgan & Claypool, 2011.

Hoekstra, R., Meroño-Peñuela, A., Rijpma, A., Zijdeman, R., Ashkpour, A., Dentler, K., Zandhuis, I., Rietveld, L. "The dataLegend Ecosystem for Historical Statistics". Journal of Web Semantics: Science, Services and Agents on the World Wide Web, volume 50, pp. 49–61. Elsevier, 2018.

van Leeuwen, M., Maas, I., Miles, A. HISCO: Historical International Standard Classification of Occupations. Leuven University Press, 2002.

van der Meer, A., Boonstra, O. Repertorium van Nederlandse Gemeenten, 1812–2006, waaraan toegevoegd de Amsterdamse code. DANS Data Guide 2, The Hague, 2006.

Meroño-Peñuela, A. LSD Dimensions: Use and Reuse of Linked Statistical Data. In: Proceedings of the 19th International Conference on Knowledge Engineering and Knowledge Management, EKAW 2014. LNCS 8982, Springer. Linköping, Sweden, 2014.

Meroño-Peñuela, A., Ashkpour, A., Guéret, C., & Schlobach, S. (2015). CEDAR: The Dutch Historical Censuses as Linked Open Data. Semantic Web – Interoperability, Usability, Applicability. 8(2), pp. 297–310, 2016.

Meroño-Peñuela, A., Hoekstra, R. Automatic Query-centric API for Routine Access to Linked Data. In: The Semantic Web – ISWC 2017, 16th International Semantic Web Conference. Lecture Notes in Computer Science, vol 10587, pp. 334–339, 2017.

Schmidt, A., van Nederveen Meerkerk, E. Reconsidering The "Firstmale-Breadwinner Economy": Women's Labor Force Participation in the Netherlands, 1600–1900. Feminist Economics, 18(4), pp. 69–96. Routledge, 2012.

Zandhuis, I., den Engelse, M., & Mac Gillavry, E. Dutch historical toponyms in the Semantic Web. In: Population Reconstruction (pp. 23–41). Springer, 2015.