

Demo: GeoDataWizard for Linked Spatial Data Creation

Alexandra Rowland¹, Jorrit Overeem² & Erwin Folmer²

¹ Kadaster & University of Twente, 7500 AE Enschede, The Netherlands
lexi.rowland@kadaster.nl

² Kadaster, 7311KZ Apeldoorn, The Netherlands
jorrit.overeem@kadaster.nl

³ Kadaster & University of Twente, 7500 AE Enschede, The Netherlands
erwin.folmer@kadaster.nl

Abstract. In order to assist users with the transformation and publication of spatial linked data on a small scale, the GeoDataWizard tool was developed by Kadaster, the Dutch Land Registry and Mapping Agency, as an open source project. The tool allows for the transformation of relational data, in a CSV format, to spatial linked data. The results of this transformation can be downloaded or published to the Platform Linked Data Netherlands (PLDN) triple store. The intention of this paper is to support users in making use of this tool.

Keywords: linked spatial data, linked data tooling, open source tooling.

1 Introduction

This paper demonstrates the GeoDataWizard tool developed by Kadaster, the Dutch Land Registry. You can find the tool as a demonstrator¹ and through a GitHub repository². The tool is an extension of the first version of the LDWizard, an open source project initiated by Network Digital Heritage³ (in Dutch: Netwerk Digitaal Erfgoed). This project resulted in a product which allows small tabular datasets (CSV files) to be transformed into linked data and the extension demonstrated here ensures that geographically related elements in a datasets, such as co-ordinates or an address, are also transformed correctly as linked spatial data. The GeoDataWizard is also available as open source software.

A unique feature of this tool is the ability to directly publish data to the triple store maintained by Platform Linked Data Netherlands (PLDN). This makes data accessible for (SPARQL) querying and visualisation in the triple store itself, but also for reuse in other applications as an endpoint. At this moment, it is possible to make relations

Copyright ©2021 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

¹ <https://labs.kadaster.nl/demonstrators/geodatawizard/#1>

² <https://github.com/netwerk-digitaal-erfgoed/LDWizard-Core>

³ <https://github.com/netwerk-digitaal-erfgoed/LDWizard-ErfgoedWizard>

between Dutch city names contained in a csv file, the Base Registry for Addresses and Buildings (Dutch acronym: BAG)⁴ and Topography (Dutch acronym: BRT)⁵.

The overall vision for the development of this tool is to assist users with the creation of spatial linked data. The publication of this newly-created spatial linked data to the PLDN triple store in particular then provides the user with multiple out-of-the-box techniques to perform analysis and visualization tasks on this data. Future developments could see this tool as a beginning tool in an overall tooling workflow which sees the csv geodata transformed to spatial linked data, uploaded to the triple store for analysis purposes and the results then visualized in a geographical viewer tool such as the toponamenzoeker⁶, also developed and maintained by Kadaster. In this way, the tool makes spatial linked data more accessible to a wider range of user groups and contexts. Future iterations of this tool have two additions planned, namely; an address matching functionality between addresses found in the csv and those registered in the BAG dataset and the automatic generation of SPARQL queries using the uploaded data; again supporting skillsets from a wider range of user groups.

2 GeoDataWizard Demonstration: A Step-by-Step Guide

2.1 Step 1: Upload the Dataset

To upload a file, click *load your CSV file* button on the main page. Currently, the Wizard only accepts CSV files and if you do not have a suitable one, there is an example csv available just below the **load your CSV file** button.

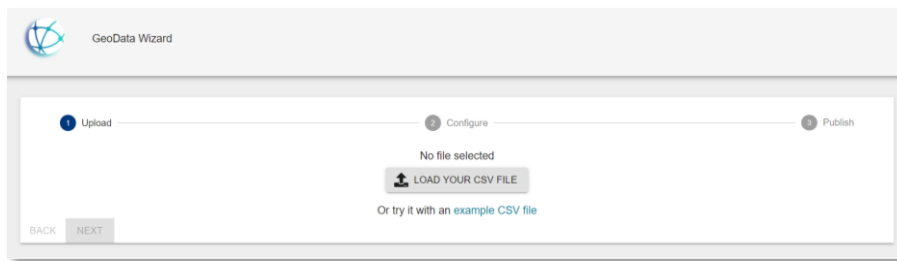


Fig. 1. Start page of the GeoDataWizard displaying the **load your CSV file** button.

Following this, a pop-up will appear asking you to choose the file you would like to upload. Select the relevant CSV file on your local drive and then click open. The window will disappear and you will be automatically redirected to the configuration step. When the CSV file is loaded, the first 10 lines of your dataset will appear in a table.

⁴ <https://bag.basisregistraties.overheid.nl/>

⁵ <https://brt.basisregistraties.overheid.nl/>

⁶ <https://labs.kadaster.nl/demonstrators/namen-app/#/>

2.2 Step 2: Configuration

Two different configurations are possible with this GeoDataWizard when a header is clicked; either configuration based on the key column and resource class IRI where these apply to the entire table (option A) or separate configurations per column (option B).

A. Key Column Configuration

You can set a key column based on which the configuration applies. These column values are added to the resource class IRI with an ID and must be a column with a unique value. When you select resource class IRI, you can set the resource IRI that applies as the relevant resource for the properties in concern. If you leave the key column and resource class empty, the default values will apply.

B. Column Configuration

Each column can be configured separately with a number of settings. These settings are important to be able to make good linked spatial data.

Datatype Setup

The datatype setup applied to your data should be done in the context of the types of analysis you are likely to perform after transformation. There are several options:

1. **String:** this is for text. This is the default value if the data type input is left empty.
2. **Integer (int):** can be used for integer data types.
3. **Float:** can be used for numbers with decimals, such as a coordinate point
4. **WKT Literal:** this can be used if coordinates are used in a separate column. Note: these values must be indicated as POINT (lat, long) in your dataset.

Property Settings

Each column value has its own property or properties in the linked (spatial) data structure. It indicates the type of value in the input. Here, a type can be specifically assigned per column.

Value Configuration Settings

With this option, you have a number of input values, all of which are used to transform the column values into IRIs. The options for this are as follows:

1. **IRI Prefix transformation:** converts the column values to an IRI as set in the resource class in the combination with the property.
2. **Search for cities in BAG:** this option is only possible for columns containing values with a city or village. Naturally, this will only work for Dutch data.
3. **Link GeoPoint:** this option is only possible if the column contains coordinate points with a POINT (...) value. This option will link the points with a BRT area in the BRT by means of an identification number.
4. **Search for places in BRT:** this option is possible if you have an address column with a street name and house number and also have a place of residence or places of birth/death in the dataset. If selected, a new select box will appear with a choice

of column names to which you want to link the address. Naturally, this will only work for Dutch data.

Once the desired configuration is set, you can press *confirm* and *next* to proceed to the publishing screen.

2.3 Publication

For the publication of data to the triple store, you do need an account to be able to request the token. If you are an existing TriplyDB⁷ user, you do not need to request a new token, you can simply input your existing one. You can request this access by contacting Erwin Folmer (erwin.folmer@kadaster.nl) where necessary. Alternatively, you can download the data in three formats, namely; CSV, RDF, and a script through which the transformation can be run manually.

Once you have access to the triple store, go back to the GeoDataWizard publication page and below the *token* input you will see 'Create a new token at: Kadaster or PLDN'. Right click on *PLDN* to open a new window or tab. You will be taken to the login page for the PLDN triple store, input you username and password. You will then be redirected to the following page (Figure 2).

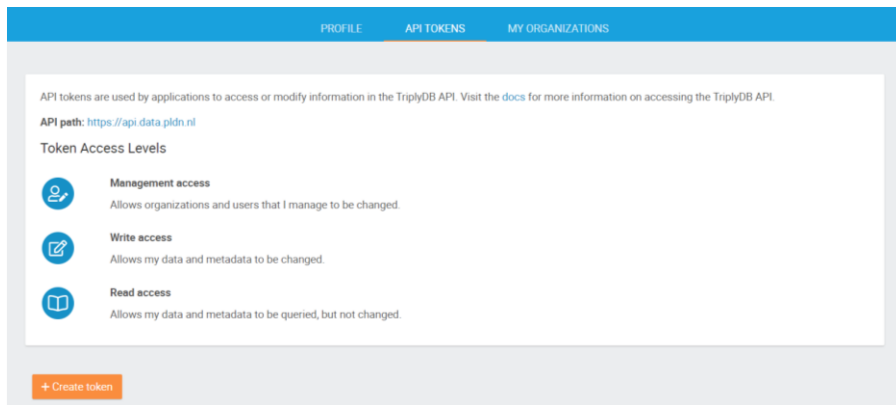


Fig. 2. Generation of an access token for the PLDN triple store.

Click *+create token*, input a token name and set the management access. A token works on three access levels and allows you to restrict your published data for other users. As a standard implementation, there is a read access restriction, with which you data can only be read but not edited by external users. Click create and your new token will appear as a popup. Copy this token and store it for future use as this is only issued once.

⁷ <https://tripllyDB.com>

After copying, click *close* and return to the GeoDataWizard and paste in the *token* in the field. Click *load token* and then the GeoDataWizard will display your account to which the dataset will be published. Click *publish* to publish your dataset. You can view now the results on the PLDN data platform through the *click here* link.