

An automated approach to ontology-based
annotation of spatial data resources on the
web with geospatial concepts

Mathijs Nelemans



Date: 2020-06-02
Supervisor: Simon Scheider
Professor: Stan Geertman

0 Contents

1 Introduction	3
1.1 Research problem	3
1.2 Research objectives	5
1.3 Scope	6
1.4 Relevance	6
2 State of the art	7
2.1 Related work	7
2.2 Spatial core concepts	7
2.3 Ontology design	10
2.4 Information extraction	11
2.5 Annotation	11
2.6 Identification	12
2.7 Information retrieval	13
3 Methodology	14
3.1 Conceptual model	14
3.2 Annotation framework	15
3.2.1 Metadata extraction	15
3.2.2 Attribute identification	17
3.2.3 Annotation rules	20
3.3 Input datasets	24
4 Evaluation	25
5 Discussion	29
6 Conclusion	31
7 References	34
Appendix A	38
Appendix B	40

1 Introduction

1.1 Research problem

The enormous amount of spatial data available online makes it difficult to use effectively. Each data source has its own unique attributes that may or may not be suitable for a particular use case. This heterogeneity of data formats is a widespread problem and requires the laborious task of integrating the data before it can be used for analytical purposes. Data integration is originally done by an expert such as a geospatial analyst by manually extracting spatial data from a large number of online resources. Next, the data is cleaned and interpreted by the expert, frequently fixing data formatting issues or matching attributes for similar data across multiple sources.

In order to retrieve the data, there are many different forms of distributions and spatial data formats. In the best case scenario, the data format for a spatial data source is implemented according to the OGC standards¹. This includes, but is not limited to, common data formats such as GeoJSON and GeoPackage, which are often retrieved using a direct download link. Other forms of distributions include web tables, SDIs, and APIs.

The use of geodata standards enables syntactic interoperability, but does not guarantee semantic interoperability. The problem of semantic heterogeneity occurs when there is a lack of contextual information, which causes confusion around the intended meaning of an attribute (Nowak et al., 2005). This could include different meaning of attributes with the same name, similar meaning of attributes with a different name, or any combination of the two (across multiple data sources).

The effective retrieval and integration of geographic information is a key factor in decision making for a variety of domains (Lutz & Klien, 2006). As the total number of spatial data sources grows, the process of manually extracting and verifying data from the web becomes more cumbersome. An expert may examine a web table to understand how values are linked and perform an analysis based on the results, but the complexity of relating a large number of tables manually makes larger tasks nearly impossible (Cruz, Ganesh & Mirrezaei, 2013).

The process of integrating spatial datasets is very similar to non-spatial datasets when just looking at the data attributes, but spatial data has an entire extra dimension that requires additional work. Spatial data usually has some type of geometry representing the location stored in the row. It may be possible that the place is only represented using text, for example using a place name. Geometry extraction and identification are crucial for spatial data, further complicating the task of data integration.

1 <http://www.ogc.org/docs/is/>

In contrast to the manual tasks, programs can be created to automate the process of data extraction and integration. Fields such as data mining and information retrieval are tasked with solving problems related to identifying relevant information from data. However, many resources are manually created and their purpose is to be interpreted by humans, not by machines. The semantic web vision aims to solve exactly this problem by creating a linked web of data, annotated with metadata that is machine readable (Berners-Lee et al., 2001). There are many benefits to this vision, but the most crucial part of the semantic web that helps to solve the problem of data integration, is the ability to enable semantic interoperability through the RDF data model and OWL (Tekli et al., 2013).

A semantic annotation is used to describe and formally identify a resource through the use of concepts and their links, often represented in an ontology (Macário & Medeiros, 2009). These annotations (and their links) will enable a program to know what to do with the resources that can be found on the web. It stands to reason that in order to enable automated data integration, online spatial resources have to be semantically annotated first.

Existing work (GeoKnow, 2019; Knoblock et al., 2012) in automated integration and annotation of datasets, often aims to implement domain specific ontologies. This research attempts to create a support system for spatial core concept annotation, which promises to bring the geospatial dimension to any discipline.

In the context of this research, online spatial resources will be linked with the *Core Concept Ontology* (Scheider et al., in-press). This ontology is developed to help answer geo-analytical questions through the generation of workflows. Geospatial data sources are the first node in such a workflow and thus the aim of this research is twofold: First to develop a method of automating the extraction and annotation of online spatial resources, and secondly to test how well data sources can be automatically annotated with the *Core Concept Ontology*.

1.2 Research objectives

The aim of this research is to develop a method to automate the extraction and annotation of online spatial resources. To achieve this purpose, the data is annotated with the *Core Concept Ontology* developed by Scheider et al. (in-press), which enables tools that are annotated with the same ontology to make use of the data. In order to test the method, a predefined set of data sources will be explored in depth by extracting and annotating them. The result of this process is then validated to ensure the successful annotation of the data. Resulting from these objectives, the main research question can be defined as:

To what extent can automated processes find, extract, and annotate spatial data from the web with geospatial concepts?

In order to answer the main research question and objectives, the following sub-questions have been formulated:

1. What are the different means of storage for online spatial datasets and in what manner can the data be extracted?

Data can be found in many different forms on the web. Data is most often exposed via human-oriented web pages that contains non-structured information about the data source. How can this non-structured information be extracted and converted into metadata to enrich the source information? And in what format can spatial data be downloaded?

2. How can attributes be automatically identified and annotated?

Data attribute headers are often fuzzy and non-machine readable. However, when combined with (statistical) information about the data instances themselves, a unique fingerprint of the attribute can be created. This fingerprint can be used to match to an existing attribute. The quality of the process behind creating the fingerprint is of high importance: the fingerprint may wrongfully assign similar values to different groupings, or cause too many conflicts and merge attributes with others they do not belong to.

3. How can geometries be matched to spatial repositories?

It is important that the geometries in a dataset are uniquely identified so that the correct spatial information can be linked to the data source. For example, identifying that a data set has data linked to municipality boundaries adds valuable information that can be used when annotating the data.

4. How can automatic annotations of geospatial concepts be validated?

Annotated datasets that have been created automatically need to be of high precision. The automated annotation process is validated against a gold standard dataset to evaluate this process.

1.3 Scope

This research is about developing a method for automating the process of extraction and annotation of online spatial data. A list of open geodata sources from the municipality of Amsterdam will be used to simulate a real world scenario. It is outside of the scope of this research to integrate this data completely, rather this study offers a proof-of-concept meant to validate the proposed methodology and demonstrate the feasibility of the automatic approach over the manual approach.

Additionally, this research is not meant to result in a repository of linked geospatial data. Indeed, the work can result in such a repository to help answer part of the research questions but the repository itself is not part of the research objectives. Hopefully, the method developed for this research project can assist future work in creating a fully functional linked geospatial data portal that can simplify data integration.

1.4 Relevance

The goal of this research is to enable the creation of automated pipelines for annotating geospatial data. The results of such a pipeline will lower the amount of time and knowledge required to integrate geospatial data in different types of analyses, opening up the wealth of geographic data to other fields and sectors.

As the result of the creation of such a pipeline, the research will further develop the standardization of spatial dataset metadata. This includes identifying existing problems in online spatial datasets and proposing solutions for retrieving metadata from modern web pages. Finally, geospatial data is annotated with core concept data types, which contain semantic information that can be used in automated geocomputations, contributing to the conceptualization of automated geospatial analysis.

2 State of the art

2.1 Related work

There is a decent amount of related work in integrating geospatial datasets (GeoKnow, 2019; Datalift, 2019; LinkedGeoData, 2019). In general, most projects start from a data source such as a CSV file or a shapefile. From the source several steps are executed on the data in order to enrich it with semantic annotations. The ontologies that the spatial datasets are linked to are generally a combination of well-known public ontologies such as GeoNames² and application specific ontologies (Prudhomme et al., 2019). The steps to annotate the data may differ from project to project because of slightly different requirements, but overall, much overlap exists.

A strong precedence in semantic type inference was set by Knoblock et al. (2012) with their semi-automated semantic type assignment to data columns. It operates on an active learning principle where a user will assign semantic labels to columns that are then used to label the header and column values. Features are constructed and extracted from the labelled data and used to train a model that can later predict unseen data. The character of the labelling system is to assign semantic meaning to the data columns to allow data integration through a domain specific ontology.

Existing spatial (GeoKnow, 2019; Datalift, 2019; LinkedGeoData, 2019) and non-spatial (Knoblock et al., 2012) semantic integration systems attempt to create domain ontologies for datasets. In contrast to these models, this research attempts to match datasets to the existing *Core Concept Ontology*, which only adds semantics to the spatial layer. Various elements of data processing techniques and attribute identification are explored and considered, and ultimately applied to annotating core concept data types.

2.2 Spatial core concepts

The geospatial dimension of datasets is not unique to a single field but is used across multiple disciplines. Kuhn (2012) defined 10 core concepts of spatial information, which are also meaningful to scientists that are not necessarily specialists of spatial information. Scheider et al. (in-press) formalizes these core concepts with common geodata types in an ontology. An important distinction is made between core concepts and data types, with the former representing human understanding and the latter representing a direct statement about the data. Therefore, the purpose of the *Core Concept Ontology* is to capture information about data types that can formally instruct spatial functions, which would normally be implicitly interpreted through the core concepts by the human analyst.

² <http://www.geonames.org/>

The spatial core concept data types (CCD) can be distinguished in three high level classes: geometric properties, core concept properties, and measurement levels. Geometric properties identify the geometry data type of a dataset, consisting of the common *Point*, *Line*, and *Region* types as well as *Tessellations*, defined in this context as a region that covers the entire extent while having no overlapping polygons. Each of these geometry types are grouped as a collection of data instances with similar geometry references under a class called *SpatialDataSet*. Identifying the geometric data type of a dataset allows an interpreter to narrow down the set of possible operations. For example, a *PointDataSet* cannot directly be used with raster operations (even though converting between the two formats is trivial).

On its own, geometric data types will not provide meaningful information for semantic interoperability. Instead, core concept properties cover the thematic information about the content of the data. These properties are based on the work of Kuhn (2012) and consist of the core concepts field, object, event, and network. Each of the core concepts have intrinsic qualities that apply to them uniquely. This distinction is where the value is added from annotating datasets with the core concept properties, as it transfers the implicit human knowledge to operational constraints to be used by the appropriate tools.

The differences between field, object, event, and network can not always be captured using rules, as they are meant to represent thematic content. For example, a field does not necessarily have to be a *RegionDataSet*, as it may also be a collection of measurements used to estimate the quality of a field using interpolation. Similarly, an object does not have to be a *PointDataSet* but could also be represented using any of the other geometric data types. Each of the thematic differences describe the qualities of an attribute of a dataset, but because core concept properties are often duplicated among the same dataset it sometimes easier to describe the dataset as a whole.

Finally, there are important attribute distinctions for geocomputations. Measurement levels (nominal, ordinal, interval, and ratio) (Chrisman, 1998) describe the types of numerical operations that can be applied to data. Within geographic applications there is another important distinction that changes the way the data can be combined: data can either be intensive or extensive (Scheider & Huisjes, 2019). Intensive measures describe a property while being independent of its size (e.g. percentage of crimes), while extensive measures are additive based on the geometries' volume (e.g. number of crimes).

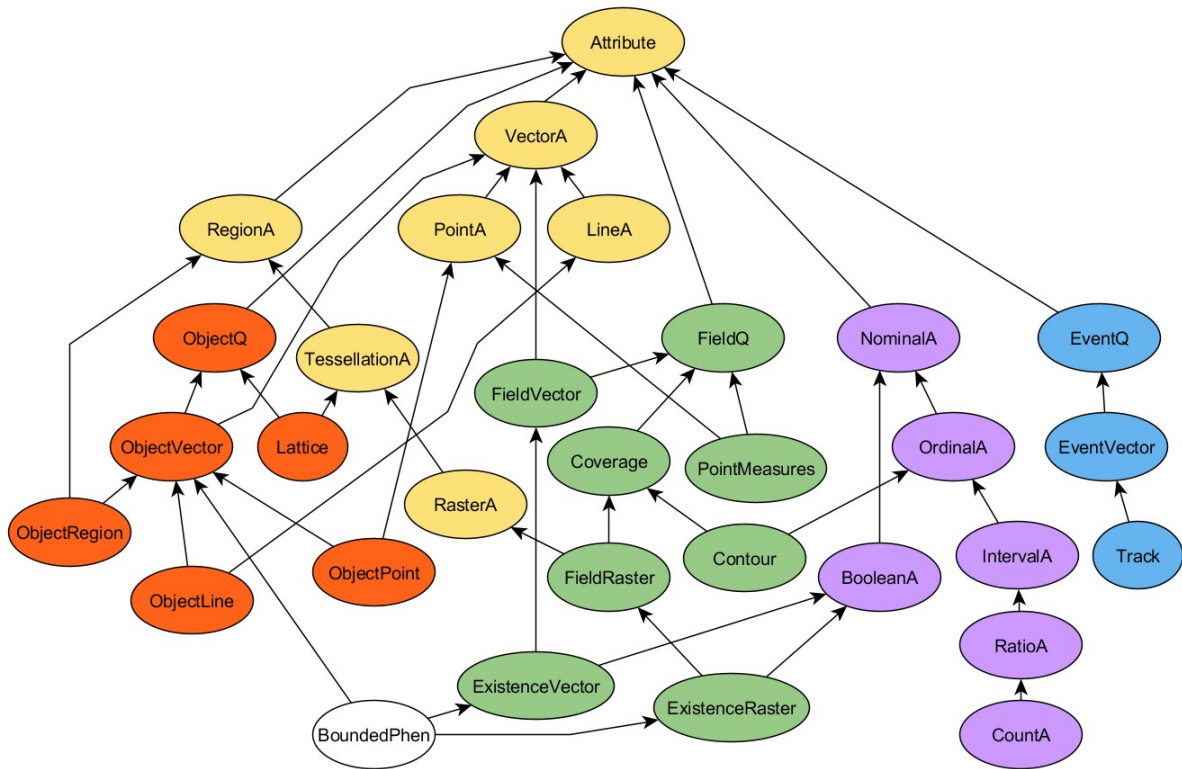


Figure 2.1: Hierarchy of CCD attribute types, showing the combination of geometric types, core concepts, and measurement levels. Yellow classes show geometric types; the green, red, and blue classes stand for the core concepts field, object, and event, resp.; violet classes denote levels of measurement. From Scheider et al. (in-press).

The CCD attribute types (figure 2.1) can be mixed freely, but certain combinations are more frequent. For example, a dataset that only contains points (*PointA*) and the object quality (*ObjectQ*) depicts a collection of real world objects such as trees. A dataset that contains objects referenced by a tessellation, may depict administrative boundaries. Because geometric and core concept data types are often the same across an entire dataset, it is preferable to simply talk about dataset qualities as the attribute qualities can always be inferred using class subsumption. The previous examples could in this case be annotated as an *ObjectDS* for the collection of trees and *LatticeDS* for the administrative boundaries. For simplicity, geometric and core concept types are assumed at the dataset level for this research.

2.3 Ontology design

In order to expand on the implementation of the *Core Concept Ontology*, this section will introduce a small primer on (web) ontologies to elaborate on the topic. At its core, ontologies group concepts into hierarchies to be reused with other ontologies. There are different levels of complexity that can be adopted, depending on the requirements of the domain. Some ontologies exclusively specify class hierarchy, often called taxonomies instead, while more complex implementations add constraints to properties and class relationships.

The most simple models can be created using RDF Schema, a vocabulary for describing resources and their relationships in RDF, with the purpose of enabling users to create basic ontologies (Brickley & Guha, 2014). More complex ontologies can utilize Web Ontology Language (OWL), which formalizes constraints of the relationships between classes. This enables reasoners to validate and infer relationships based on a subset of rules (Horrocks, Patel-Schneider & Van Harmelen, 2003). The *Core Concept Ontology* is implemented in OWL, specifying both class hierarchy and restrictions on relations such as a `PointDataSet` being disjoint with a `RegionDataSet`.

There are four types of ontologies: top-level, domain, task, and application ontologies (Guarino, 1997). Ontologies are meant to be reusable models, but as you move from top-level to application ontologies, the implementation becomes less generic and reusability suffers. Highly adapted top-level implementations such as FOAF and SKOS are able to support almost any domain (Presutti & Gangemi, 2008), but do not provide enough constraints to help with the execution of application logic.

The *Core Concept Ontology* aims to implement just enough properties to allow the execution of functions on annotated datasets, while retaining the ability to apply the core concepts to multiple disciplines. In order to retain this generic overview, other high level models such as SKOS are a good match with the spatial core concepts. SKOS³ is a simple representation that labels entities as concepts, that aggregate into concept schemes. Note that the term concept in relation to SKOS represents any type of meaning or category, purposefully kept as abstractly as possible. In order to avoid confusion with the spatial core concepts, SKOS concepts will from here on always be referenced by prefixing “SKOS”.

2.4 Information extraction

Information extraction is the task of automating the translation of unstructured content into structured information. Translating unstructured information about datasets from web pages into a structured form is a type of information extraction.

Documents can either be unstructured, semi-structured or structured, which generally relates to free text, HTML, and databases, respectively. Programs that execute the task of information extraction are called wrappers, and they act as a translation layer between the source document and the target query (Chang et al., 2006). Wrappers are commonly used to extract information from two types of online resources: web pages (HTML) and data APIs. Extracting content from web pages is more flexible because it can be used on most pages, but APIs are easier to access. Both types of resources can change at any time, for example, with the release of a new page layout or a new version of the API, breaking the wrapper and rendering the time invested to create it, useless (Freelon, 2018).

³ <https://www.w3.org/TR/skos-primer/>

There are two methods of information extraction from the web (web scraping): coding a custom wrapper and utilizing online services. In order to program a custom scraper, a good understanding of programming, HTML, and the Document Object Model (DOM) is required (Krijnen, Bot & Lampropoulos, 2004). The required levels of understanding are much reduced when using an online service such as a browser extension.

Web pages can be categorized as semi-structured data because the components that make up a web page are stored hierarchically in the DOM-tree (Ferrara et al., 2014). This feature of HTML pages is used to access and extract data from web pages in a technique called Xpath selection. Xpath is a language that describes the location of an HTML tag within the DOM-tree. When utilizing Xpath selection, the wrapper will locate the node and extract the HTML tag within that node, which has the added benefit of utilizing the existing structure of HTML. Other extraction techniques have the disadvantage of not utilizing the structure present in HTML, but this may be more desirable for cases where the structure is illogical. These techniques are regular-expression (regex) and machine learning-based approaches (Ferrara et al., 2014).

Finally, some web resources already contain structured data. Integrating data that is already semantically annotated only requires linking the vocabularies, but far too little web resources are at this stage. A new type of structured data that is showing up on an increasingly growing number of websites are markup standards such as Microformats, Microdata, and RDFa. Sites have started annotating content with these formats in order to increase their search ranking. Bizer et al. (2013) have found structured data within 369 million out of 3 billion pages (12.3%) hosted on Common Crawl, highlighting the growth of structured data. However, most of these pages are news and entertainment websites and only 63,668 websites provide useful metadata through the Dublin Core vocabulary.

2.5 Annotation

The first step of annotating a dataset is to identify what the column labels are and what they represent. Finding the meaning of a column is often achieved by looking at the data values, and not the column header itself. Automated semantic annotation can be distinguished between rule- and machine learning-based methods, with the latter requiring a training set of annotated data (Reeve & Han, 2005). Thus, pattern recognition in the data is used to identify a concept that is present in the ontology that the dataset is being annotated with. This is frequently executed on a sample of the data in order to speed up execution times, depending on the complexity of the method (Knoblock et al., 2012).

Most examples of column detection in related work identify a certain set of important, frequently occurring patterns. Prudhomme et al. (2019) identify an ID column based on an incrementing integer that has a unique constraint. Furthermore, address components, dates, urls, etc. are extracted using regex patterns. All leftovers

can be either a number (double or integer) or remaining string, which may either indicate a class type because of many recurring appearances or it indicates a description. Neumaier et al. (2018) detect geographical identifiers such as NUTS identifiers and postal codes. This is possible because these identifiers consist of very restrictive patterns.

Because pattern based annotation is only able to annotate identifiers and frequently occurring attributes (such as a postal code), this does not result in the ability of linking the data attributes to the geometry. One possibility is to use the labels that identify the data with the help of string-based similarity (Klien, 2007; Prudhomme et al., 2019). This poses some problems as column headers are often represented in natural language, and this is not guaranteed to even be in English. Prudhomme et al. (2019) propose to translate the labels and to match them with existing concept names, or to create a new concept if it does not yet exist. However, such an approach may cause problems due to different providers using different terminology, which can cause low precision when querying the data (Lutz & Klein, 2006).

More dynamic textual extraction may be achieved using a common method in full text search which uses the product of the *term frequency* and *inverse document frequency* (TF-IDF). TF-IDF indicates how unique a term is in relation to a document, which promotes unique terms and demotes common english terms (Manning, Raghavan & Schütze, 2008). In Ramnandan et al. (2015) the key terms are extracted and used to identify a semantic label by finding similar sets of key terms. Key terms can also be used in machine learning models by transforming text into so called bag of words and assigning weights based on term frequency. These text categorization models are frequently variants of Naive Bayes or Support Vector Machines as they are relatively easy to work with as a baseline method (Wang & Manning, 2012).

2.6 Identification

After linking data attributes to a geometry, it is beneficial to enrich the geometry itself as well. Just linking values to a geometry will allow geospatial queries to extract information from them, but it will cause problems with duplicate data or when searching for a place name. Related research (Zhang et al., 2013; Ateazing, 2015) demonstrates some basic functions to determine if two geometries are the same. For points, simply taking the euclidean distance is enough to determine spatial proximity, but as there is only one point of verification this may also cause false positives. For complex geometries the Hausdorff distance returns the greatest distance between two geometries, which should be near zero for identical linestrings or polygons.

Linking geometries to a central repository such as GeoNames or OpenStreetMap ensures that administrative boundaries can be matched between datasets and linked to a pre-existing high quality database (Neumaier et al., 2018; Prudhomme et al., 2017). The GeoNames ontology contains the hierarchical relations between administrative boundaries, and the class names describe the type of place (e.g. city or country). Linking the dataset geometries to GeoNames or a similar repository would

therefore automatically add additional metadata, such as what kind of geographical area the dataset is describing.

With the help of a places repository, querying the extent of a dataset becomes much more intuitive. People tend to use natural language to describe their intentions and places are no exception (Montello, 2003). Capturing the spatial extent of a data source is trivial when describing it with a geometry, but linking geometries to place names allows future users to describe their spatial extent with words rather than coordinates.

2.7 Information retrieval

Geographic Information Retrieval (GIR) is an extension to the field of Information Retrieval (IR). Because an immense amount of data is geographical in nature, while also mostly unstructured in the form of free-text, GIR concerns itself with combining the capabilities of structured spatial filtering with the unstructured key term searching methods of IR (Jones & Purves, 2008).

There are a couple of issues in GIR, most notably related to identifying the geographical reference because of natural language and name ambiguity. Next, ranking the relevance of documents is complicated because of the combination of spatial and theme-based search terms. This also translates into difficulties when evaluating the success of a GIR method (Jones & Purves, 2008). Measuring the effectiveness of a retrieval revolves around identifying the relevant and non-relevant documents. In order to test a retrieval, a gold standard is created that consists of a collection of documents that is pre-ranked as either being relevant, or non-relevant (Manning, Raghavan & Schütze, 2008).

The most frequent measures for IR effectiveness are precision and recall. Precision (P) is the fraction of retrieved documents that are relevant, and recall (R) is the fraction of relevant documents that are retrieved. A single measure that trades off precision versus recall is the F measure. A balanced F measure equally weights precision and recall and is commonly written as $F_{\beta=1} = 2 \frac{PR}{P+R}$, although different notations exist when using unequal weights (Manning, Raghavan & Schütze, 2008).

Precision and recall are originally used in an information retrieval scenario where the number of relevant documents need to be measured. They can also be used in a classification scenario, where precision is the number of predicted true positives divided by all predicted positives and recall is the number of predicted true positives divided by all true positives in the gold standard. Annotating the datasets and attributes closely resembles a classification task, where precision and recall may be used to qualify the annotated data. This is used to measure the accuracy of the annotation process in the evaluation section.

3 Methodology

This research will make use of the Amsterdam open geodata portal⁴ for a collection of datasets with different geometries and themes.

The datasets hosted on the geodata portal differ in types of datasets, geometry types, and themes. The different types of datasets and the attributes present in the datasets are explored and a method is developed to annotate them (semi-) automatically. After annotating the data, geometries are compared to a spatial repository to identify existing places and shapes. Based on various rules some datasets or attributes can be automatically annotated with the appropriate label, while others need human supervision to train the annotation process.

In order to assist the human labeler, a geospatial annotation tool called MetaMapper was created for this research. The tool facilitates all the data processing steps required to annotate datasets. Attributes or datasets that cannot be automatically annotated, may instead be labeled by the user. In the case of attributes, the newly added label is added to all data instances which increases the amount of training data that the tool can use to match unseen data to existing data.

The sources have been manually annotated with the core concept data types to serve as the gold standard dataset against which the automated annotation method will be tested. The entire process from sourcing to ingestion, both manual and automatic, and the answers to the four sub-questions help answer the main research question.

⁴ https://maps.amsterdam.nl/open_geodata/

3.1 Conceptual model

Figure 3.1 represents a graphical representation of the methodology and how it relates to each of the four research questions. The conceptual model is used as a framework to structure the methods introduced in the following sections. First the different data processing elements that are required to setup the annotation system are explained. Next, the annotation rules that utilize these data processing techniques are described. Finally, the developed system and its output is tested in the evaluation.

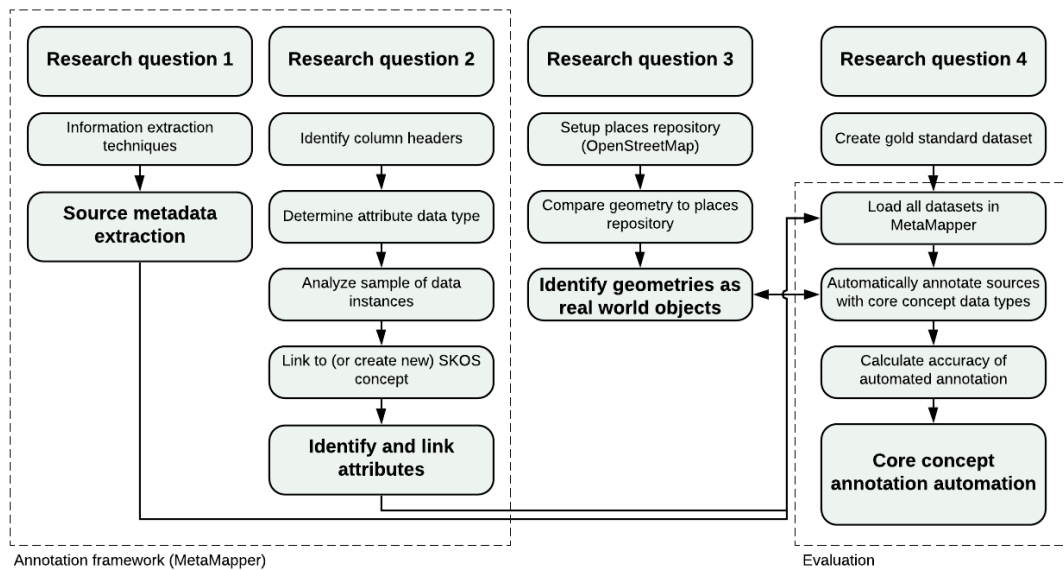


Figure 3.1: Conceptual model

3.2 Annotation framework

In the context of this research, a geospatial annotation tool called MetaMapper⁵ was developed to facilitate extraction, ingestion, identification and annotation of datasets. The tool is specifically developed for geospatial datasets and annotating them with core concept data types.

Each dataset is added to the system by adding the URL to the online webpage that serves the metadata about the dataset. This webpage is viewed as the publisher of the data source, and metadata about the dataset is then extracted by opening the webpage in a wrapper. After adding the new dataset to the system, it may be ingested to serve a sample of data and attempt the annotation of the data with the CCD ontology. Each attribute is either matched to a previously annotated attribute, or a new one is saved automatically. This increases the amount of training data to identify the type of attribute automatically, or allows inference when a user has manually

⁵ <https://github.com/quangis/metamapper>

annotated such a column. Finally, an RDF representation of the annotated datasets and their attributes can be exported.

3.2.1 Metadata extraction

Even though only one portal is used for all input datasets, most websites that host spatial datasets consist of different landing pages and content. Although they represent the same type of data it is not trivial to automatically consume the data in a similar manner. In order to be able to support any type of website, the extraction of metadata is accompanied with an information extraction wrapper that scrapes the xpath of the selected element. Using this method, metadata such as the title or the download URL may be saved based on their location instead of a static snapshot. If the wrapped elements were to be changed the metadata about this dataset can be updated automatically.

Modern web pages often use dynamically generated URLs as download links. Traversing the clicks saved using the wrapper ensures that it can always be programmatically accessed in the future. Saving clicks and metadata fields is accomplished by launching the web page in Selenium, a tool that allows programmatic access to the browser driver. Some javascript is injected into the web page to add custom functionality that can save the xpath and report it back to the application. When the data is extracted from the web page, the saved xpaths are executed in the same order as originally saved to ensure that all elements (e.g. tabs, filters, and external pages) are loaded correctly.

Box 3.1: Metadata extraction example

In order to better portray the functionality of the MetaMapper tool, an example scenario concerning a dataset containing all buildings that are part of social housing corporations in Amsterdam is annotated with core concept data types. The data is published on the Amsterdam open geodata portal⁶.

The source is added (a) to MetaMapper by submitting the access URL (b) of the dataset. The access URL locates the webpage that publishes the dataset, but does not necessarily link to the dataset itself. More metadata is required to enable the ingestion and annotation of the dataset, at minimum the download URL and a title. The metadata extraction wrapper is launched when the source is edited or updated (c), navigating to the access URL and waiting for the user to annotate the required metadata (d). During the metadata extraction, clicks are recorded and only the location of the metadata elements on the webpage are saved. When the wrapper is closed, the clicks are replayed and the actual content is extracted and saved (e).

⁶ https://maps.amsterdam.nl/open_geodata/?k=208

a)

MetaMapper

Browse existing or annotate new geospatial data sources

Add Source

Browse Concepts

b)

Submit

Cancel

c)

access_url

Create / Edit

Ingest

Annotate

Delete

d)

Privacy - vragenpanscan
Zonnepanelen op erfgoed
Percentage
Bouwjaar

Select annotation type

download-url

OK Done

Stedelijke ontwikkeling
Corporatiebezit AFWC
Functiemix
Grondgebruik 2017
Jongeren- en Student
Niet-woonfuncties (Fu
Oergeul
Oergeul - sonderingen
Ontwikkelbuurten
Overnachtingsbeleid
Stadsloods
Woningbouwplannen

dam.nl/afwc_2019/ >
tte van Woningcorporaties

U gaat akkoord met de [Gebruiksvoorwaarden](#) >
Download: [.csv \(Excel\)](#) >
Download: [.json \(GeoJSON\)](#) >

e)

AFWC_2019

access_url

download_url

title

Create / Edit Ingest Annotate Delete

3.2.2 Attribute identification

When a download URL is successfully extracted for a new dataset, it can be ingested to start the annotation process. Most datasets come archived and need to be uncompressed first, after which the main content is passed to the GDAL/OGR library. GDAL⁷ is a translation library for hundreds of raster and vector geospatial data formats. In order to support a unified query interface, all data is then loaded into a postgres database.

In order to increase the annotation capabilities, attributes are matched with each other. Attributes are defined as SKOS concepts, which allows categorization and hierarchy without enforcing any domain logic. There are two levels of SKOS concepts currently implemented, a high level SKOS concept created by human annotators and their narrower relationships that represent the raw attributes (figure 3.2).

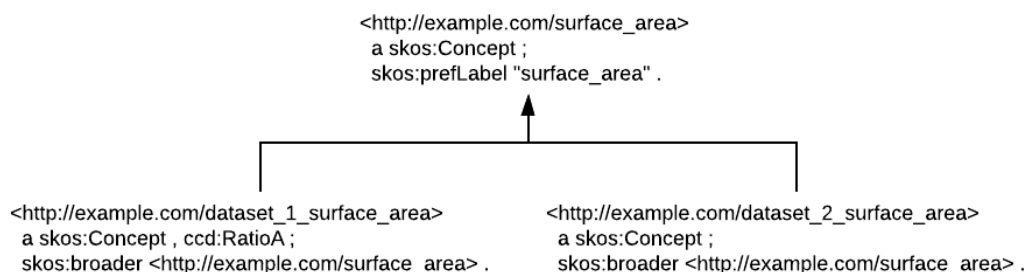


Figure 3.2: Attributes represented as SKOS concepts. The higher level SKOS concept is created by a human annotator, with two narrower (generated) SKOS concepts that represent a dataset attribute. In this example, “dataset_1_surface_area” has more annotations than the other SKOS concepts, but they are all able to inherit this information when required.

All of the data instances are used to identify similar attributes, and a new SKOS concept is generated when no existing SKOS concepts can be matched. There are two approaches to finding similar attributes, distribution comparisons for numeric data and a categorization model for textual data. Because similarity matching is not the main research objective, both types of data employ the simplest approach for identifying attributes.

For each set of numeric data instances a t-test is performed against existing SKOS concepts and their data instances. If the p-value is larger than 0.05 the null hypothesis that both distributions are similar cannot be rejected and the SKOS concept will be matched to the new data.

A Naive Bayes classifier is used to predict whether a set of textual features belongs to an existing SKOS concept. For each of the existing SKOS concepts, their data instances are labeled with their identifier and passed into a supervised text categorization model. Each instance contains a string that is vectorized and normalized using a TF-IDF transformation. These labelled features are then passed as training data into a

⁷ <https://gdal.org/>

Naive Bayes classifier, which will predict the most relevant SKOS concept for unseen data. At least half of the data instances need the same label from the predictor in order to be categorized.

When the number of existing SKOS concepts is still small, there is an issue with bootstrapping the model because it lacks negative labels (e.g. with N=1 SKOS concepts it always returns that category with a probability of 1.0). To offset this issue, a dummy category with random text snippets from news articles⁸ is added to the model, which adds random background noise that only columns with a high occurrence count (e.g. categorical data) are able to move past.

Note that the textual categorization classifier is a supervised model, but it is only supervised in the sense that the model requires labeled training data, which is provided through the linked SKOS concept. Even if no human annotation occurs and no high level SKOS concepts are generated, the classifier will still attempt to match unseen attributes with existing ones. When more than half of the unseen data has been predicted to be the same category, the new attribute is assigned to that SKOS concept as well. After each new attribute the classifier is retrained using all the existing attributes (including the new one) and their data instances, as well as a dummy category when there are less than 10 existing categories.

When two attributes are matched, it does not imply semantic similarity, this is instead achieved through the high-level SKOS concepts that are exclusively created by human annotators. Matching attributes without any semantic context is still useful for the annotation process, as two attributes with different meaning but similar data will still receive the same core concept data type. This will improve the ability of spreading human annotations across as many datasets as possible, but is clearly not a requirement for completing the annotating process itself. Therefore, the threshold of attribute identification is kept at a conservative parameter based on expert judgement, preferring to keep attributes isolated rather than matching as many as possible.

Box 3.2: Attribute identification example

Continuing with the example scenario of social housing corporations in Amsterdam, the dataset is ready for ingestion and identification. This process does not require much human intervention at all, as ingestion is possible due to the standardization of data formats and the supporting logic in GDAL, and identification is a fully automatic task at this stage.

After pressing the ingest (a) button, the dataset is downloaded and loaded into a postgres database. For each new attribute, all data instances are pulled and tested against existing SKOS concepts. For this dataset, there are a handful of columns that describe the different combinations (and their amount) of unit types in the building, a corporation name, and a construction year (b).

8 https://scikit-learn.org/stable/modules/generated/sklearn.datasets.fetch_20newsgroups.html#sklearn.datasets.fetch_20newsgroups

Going from left to right, the first textual column “*Corporatie*” is subjected to a Naive Bayes classifier, but there are no existing SKOS concepts yet. The first numeric column “*Corporatie_woningen*” is tested against existing numeric attributes using a t-test, but no such attributes exist yet. From this moment onwards, each column is tested against the previously created SKOS concepts, creating a new SKOS concept every time it does not match to another attribute. Finally, when testing “*Woningen_in_blok*” to the SKOS concept that represents “*Corporatie_woningen*”, the t-test returns a high p-value (≥ 0.05) and both attributes are merged.

Even though the two attributes are semantically different, they contain the same data because they both count the number of units in a building, although one counts the number of units owned by the corporation and the other the total. However, in most cases the corporation owns the entire building and the two attributes are nearly identical.

At this stage nothing happens yet, and the process continues for the remaining attributes. The two attributes that were matched are now linked through their SKOS concept, which means that as soon as just one receives the appropriate CCD annotation (*RatioA*), the other will inherit this as well (see also: figure 3.2).

a)

AFWC_2019

access_url

download_url

title

Create / Edit
Ingest
Annotate
Delete

b)

Corporatie	Corporatie_woningen	Zelfstandig	Onzelfstandig	Woningen_in_blok	Percentage	Bouwjaar	CORP	PERC
de Alliantie	44	44	0	44	100	1005	AL	100
de Alliantie	48	48	0	48	100	1005	AL	100
de Alliantie	35	35	0	35	100	1005	AL	100
de Alliantie	12	12	0	0	0	1005	AL	999
de Alliantie	9	9	0	9	100	1005	AL	100
de Alliantie	3	3	0	5	60	1005	AL	50
de Alliantie	6	6	0	6	100	1005	AL	100
de Alliantie	5	5	0	5	100	1005	AL	100
de Alliantie	5	0	5	0	0	1005	AL	999
de Alliantie	5	5	0	5	100	1005	AL	100

3.2.3 Annotation rules

There are two processes that may annotate a core concept data type to a dataset or attribute. First, a user is able to add new SKOS concepts to the system that supersedes the auto generated ones. Secondly, all of the data instances of the (merged) attributes are accessed to extract properties that help to identify a core concept data type. Not all data types can be annotated automatically because the system lacks the semantic knowledge for some of the assumptions, which is where the high-level SKOS concepts created by users can take over instead.

Manually creating a new SKOS concept is simplified by showing a sample of a dataset with the ability to choose an existing or create a new SKOS concept for a column. The user generated SKOS concepts are added as a broader class so that they may encompass multiple attributes over multiple datasets without trouble. Human annotation of the core concept data types is enabled with a simple dropdown from which the appropriate data type for a high-level SKOS concept can be selected.

Without manual intervention, there are a limited set of rules that can still assure the correct annotation of a core concept data type. Most notably, classes of the geometry types can generally be annotated by inspecting the geometry type in the dataset. The core concept properties and measurement levels have mixed results, because not all requirements for a data type can be guaranteed without any semantic context.

Annotation type	Decision rule
PointDataSet	Dataset only contains points
LineDataSet	Dataset only contains lines
RegionDataSet	Dataset only contains (multi-)polygons
VectorTessellation	Dataset is a <i>RegionDataSet</i> , has no gaps and no overlap
ObjectDS	20% of dataset has a Hausdorff distance of 15 meters or lower to any object in OpenStreetMap
LatticeDS	Is both a <i>VectorTessellation</i> and <i>ObjectDS</i>
NominalA	1-20 unique instances
BooleanA	Is <i>NominalA</i> and exactly 2 unique instances

Table 3.1: CCD annotation decision rule overview. The missing CCD annotations are unable to be matched automatically.

Geometry type detection is straightforward due to the data instances being ingested into a postgres database, where the geometry types can simply be read. In case of polygons (*RegionDataSet*) there is a narrower class called *VectorTessellation* that is

not standard to the geometry types in GIS. However, the definition of a *VectorTessellation* (*RegionDataSet* that is non overlapping and covers the entire extent) can still be transformed into a query which allows programmatic detection.

Core concept properties include assumptions about the data that influences whether an operation on the data would be meaningful or not. Determining whether or not a dataset type is correct usually requires external knowledge about the intent of the data. However, there are a few common dataset types that can usually be guaranteed to be correct when detected. These include the *FieldRasterDS* in case of rasters, and in case of vectors both *ObjectDS* and *LatticeDS*.

Detecting whether a dataset contains objects is achieved by comparing the geometries with a centralized repository of places and regions. This repository consists of the entire OpenStreetMap database (filtered for the city of Amsterdam in this research). Comparisons are based on the Hausdorff distance. If the distance is very small the geometries are identical. The distance cutoff (15 meters) is chosen by taking the median distance after comparing geometries that are known to represent the same area, rounded up.

When the majority of a dataset is similar to the geometry of a confirmed real world object, it can be annotated with at least *ObjectDS*. In case of a dataset with objects, a lattice may be inferred in the same manner as a *VectorTessellation* by confirming non overlapping coverage of the extent. This is achieved by joining all the individual polygons in the dataset and comparing the summed area with the area of a concave hull representation of the same dataset, which should be near identical for a *VectorTessellation*.

In case of the measurement levels, only (bi-)nominal data can be estimated by comparing the count of unique instances with the total and at least interval for numeric data. Both ordinal and ratio scales require human insight.

Box 3.3: Annotation example

To complete the example scenario of social housing corporations in Amsterdam, there are two possible routes. The first possibility is to export the annotated dataset directly, which will only apply the decision rules in table 3.1. If instead, the dataset needs to be annotated to the highest possible standard, manual intervention is required.

Going to the annotation overview (a) will preview a sample of the dataset and the show the autogenerated SKOS concepts in brackets. A new SKOS concept can be created and linked by clicking on the attribute name and writing an appropriate label. The new SKOS concept can now be searched and selected in the concept browser (b), and finalized by selecting the correct CCD annotations (c).

a)

	Woningen_in_blok (http://example.com/97325fd3174645d88c140266b30d6170)	Zelfstandig (t
	building_block	44
		48
35		35
0		12

b)

Search

is_lighted

surface_area

is_covered

building_block

c)

Measurement Scale

Ratio

Property Type

Extensive

DataSet Type

ObjectDS

Submit Cancel

3.3 Input datasets

The open geodata portal of Amsterdam contains a collection of spatial datasets maintained by the municipality of Amsterdam. Each dataset may be downloaded in different formats, most notably CSV, GeoJSON, Shapefile and MapInfo. Most datasets are also provided as a web map which, through manual inspection, resulted in the creation of the gold standard dataset. The entire overview can be accessed in the annex (Appendix A & B) and a summary of the annotation types in tables 3.2, 3.3, and 3.4.

Table 3.2: Geometric properties

Geometry	Count
PointDataSet	31
RegionDataSet	14
LineDataSet	7
VectorTessellation	8

Table 3.3: Core concept properties

Dataset	Count
ObjectDS	36
NetworkDS	8
CoverageDS	5
EventDS	4
ContourDS	2
LatticeDS	2
PatchDS	2
LineMeasuresDS	1
PointMeasuresDS	1

Table 3.4: Measurement levels

Dataset	Count
NominalA	110
RatioA	40
OrdinalA	25
BooleanA	17
IntervalA	2

4 Evaluation

With the help of the annotation framework, each of the input datasets have been processed by extracting their metadata, ingesting and annotating the attributes according to the decision rules in table 3.1. This section will shortly discuss the metadata extraction process, followed by the creation of the gold standard dataset using the annotation framework, and finally the automated annotation accuracy is measured against the gold standard.

Metadata extraction

In order to load each of the input datasets, all the access URLs have been loaded into MetaMapper and the source metadata has been extracted for each data source. This process was repeated for each data source, even though they all originated from the same webpage. The resulting database contains the required page navigations and xpaths to be able to refresh the source metadata when desired.

Creation of gold standard dataset

The data processing and annotation steps are verified by comparing the annotated datasets and attributes to the gold standard. The gold standard consists of all datasets fully annotated by an experienced (human) annotator. The gold standard was partially created using MetaMapper to simultaneously evaluate the manual annotation process. Both the gold standard and automated annotation results were created in separate environments.

To evaluate the process of creating a high quality annotated dataset such as the gold standard dataset, a subset of 20 input datasets was annotated using both MetaMapper and a manual approach. The step by step overview of the manual approach consists of navigating to the data publisher, downloading the dataset, loading the dataset in QGIS, inspecting the data using the attribute table, and finally saving the column header and its corresponding core concept data type in a spreadsheet.

Title	Download URL	Duration (min)	
		Manual	MetaMapper
AFWC_2019	https://maps.amsterdam.nl/open_geodata/?k=208	3.1	2.4
GEBIEDEN22	https://maps.amsterdam.nl/open_geodata/?k=204	2.7	1.8
VOGELS	https://maps.amsterdam.nl/open_geodata/?k=87	4.7	5.0
BUURTCOMPOST	https://maps.amsterdam.nl/open_geodata/?k=228	2.0	1.9
ENERGIE_WINDVISIE	https://maps.amsterdam.nl/open_geodata/?k=96	1.7	1.0
OPENBARE_TOILETTEN	https://maps.amsterdam.nl/open_geodata/?k=240	1.5	2.1
VERKEERSLICHTEN	https://maps.amsterdam.nl/open_geodata/?k=61	2.2	2.9
ECOPASSAGES	https://maps.amsterdam.nl/open_geodata/?k=63	4.2	2.4

BBQ	https://maps.amsterdam.nl/open_geodata/?k=319	2.3	1.7
WONINGWAARDE_2018	https://maps.amsterdam.nl/open_geodata/?k=52	1.3	0.9
GELUID_VERKEER	https://maps.amsterdam.nl/open_geodata/?k=267	3.1	0.7
SPORT_OPENBAAR	https://maps.amsterdam.nl/open_geodata/?k=56	6.1	3.3
WALKABILITY	https://maps.amsterdam.nl/open_geodata/?k=298	4.6	2.0
OERGEUL	https://maps.amsterdam.nl/open_geodata/?k=242	0.9	1.2
EPR_NESTEN	https://maps.amsterdam.nl/open_geodata/?k=67	0.7	1.5
CANADESE_POPULIEREN	https://maps.amsterdam.nl/open_geodata/?k=232	3.7	2.1
MARKTEN	https://maps.amsterdam.nl/open_geodata/?k=123	2.3	4.2
MONUMENTEN	https://maps.amsterdam.nl/open_geodata/?k=155	5.4	2.4
ZONNEPANELEN2017	https://maps.amsterdam.nl/open_geodata/?k=251	1.4	1.1
FUNCTIEKAART	https://maps.amsterdam.nl/open_geodata/?k=49	4.5	2.4
Average:			
		2.9	2.2

Table 4.1: Comparison between manual dataset annotation and MetaMapper annotation durations.

Table 4.1 shows the duration in minutes for both the manual and MetaMapper approach to annotating the gold standard dataset. The average duration for the manual approach is 2.9 minutes per dataset, and the average duration of the annotation using MetaMapper is 2.2 minutes per dataset. As is clear from the data, using MetaMapper is slightly (26%) faster. Both approaches created the same dataset as the automated annotation in MetaMapper was overruled for all attributes by creating new SKOS concepts for each relevant attribute.

Automated annotation accuracy

To assess the ability to automatically assign core concept data types to datasets, the automatically annotated data was compared to the gold standard dataset. Each type is validated within its own class. In order to reduce bias, no human supervision was applied by adding SKOS concepts. Choosing what attribute to correct would have significant impact on the outcome of the assessment. Only the non supervised results are reviewed to present a baseline for future improvements.

Core concept data types are part of a hierarchy and directly comparing the gold standard data type and labelled data type seems unintuitive at first, but if class subsumption is taken into account the precision of the labelling process would increase simply because of a decrease in the evaluation sensitivity. In some cases a dataset may be annotated with multiple data types of the same class (e.g. *ObjectDS* and *NetworkDS* in the case of a road network). In order to simplify the validation process, only the most granular of the data types in the gold standard is kept.

The annotation accuracy is measured using the F measure, which balances both precision and recall. For geometric and core concept properties, 60 datasets (table 3.2 and 3.3) have been annotated manually to serve as the gold standard and compared to the automatically generated results. Table 4.2 displays the precision, recall, and f-measure scores for the geometric properties.

Data type	Precision	Recall	F-measure
PointDataSet	1.00	1.00	1.00
LineDataSet	1.00	1.00	1.00
RegionDataSet	0.60	0.92	0.72
VectorTessellation	0.00	0.00	0.00
Total	0.85	0.85	0.85

Table 4.2: Geometry data type annotation results

The precision and recall, 0.85 for both, are very consistent for the geometry type annotations, which is to be expected because the type of geometry is easy to detect. Only the more precise geometry type *VectorTessellation* cannot be annotated directly and causes for the decrease in the otherwise flawless type assignment. All *VectorTessellation* datasets were assigned *RegionDataSet* which is technically correct but not precise enough, and one *RegionDataSet* was (incorrectly) assigned *VectorTessellation*.

Data type	Precision	Recall	F-measure
ObjectDS	0.81	0.36	0.50
LatticeDS	0.00	0.00	0.00
Total	0.76	0.21	0.33

Table 4.3: Core concept data type annotation results. Note that only core concepts that can be assigned using the decision rules (table 3.1) are shown, all others automatically score 0.00 (as represented in the combined F-measure).

As previously noted, the core concept properties are much more difficult to label and a strategy is employed that attempts to only annotate those data types that are very certain to be correct (table 3.1). This is apparent in the valuation results (table 4.3), with a high precision (0.76) and low recall (0.21), because a core concept data type was more often missing than wrong. Because *VectorTessellation* was incorrectly assigned in the previous step, no *ObjectDS* annotations have been upgraded to *LatticeDS*. All other core concept data types score 0.00 for all metrics because they

cannot be automatically assigned, but because slightly more than half of the datasets contain objects, the final F-measure is still 0.33.

Data type	Precision	Recall	F-measure
BooleanA	1.00	1.00	1.00
NominalA	0.72	0.51	0.56
IntervalA	0.04	1.00	0.08
Total	0.54	0.40	0.46

Table 4.4: Measurement level annotation results. Only measurement levels that can be assigned using decision rules (table 3.1) are shown, *OrdinalA* and *RatioA* automatically score 0.00.

In order to assess the measurement level annotations, all the 379 attributes of the 60 datasets have been annotated manually for the gold standard dataset. From the total of 379 attributes only 194 have been assigned an attribute type in the gold standard dataset (table 3.4). All the remaining attributes are not meaningful for geospatial analysis and do not contain any relevant computational information.

Automatic annotation is limited because only (bi-)nominal and interval measurement scales can be predicted, as the context to determine ordinal or ratio is missing from the data. Human supervision can overcome this limitation but in this evaluation all ordinal and ratio scale attributes have been predicted incorrectly. Note that any attribute that was not labeled in the gold standard because it does not contain meaningful data (e.g. descriptions, identifiers, comments, map symbols) may be labelled with an attribute instead. This form of a type 1 error is removed during the validation instead, because it does not cause real world errors as the attributes can not be used meaningfully in any workflow.

Table 4.4 shows the accuracy of the measurement level annotations. The precision and recall are 0.54 and 0.40, respectively. The recall for *IntervalA* seems extremely accurate (1.00), but this is skewed due to the low amount of actual interval attributes in the gold standard. Additionally, all numeric data is given the interval scale, as no distinction can be made between interval and ratio.

5 Discussion

The purpose of this research is to facilitate the annotation of spatial core concepts to support research in automated geographical analysis. In this context, it is not only relevant to deliver a high standard of annotated geospatial data but also to enable the annotation process for new datasets that are to be used in such an analysis. As demonstrated in this research, fully automated geospatial annotation is not very reliable. Instead, a combination of human and machine intelligence may provide a faster annotation process by suggesting what is already known by the system and leaving semantically difficult problems to the annotator. Additionally, the task of data retrieval and integration still leaves a lot to be desired, further increasing the difficulty of using datasets programmatically.

Due to the limitations of this research, only a limited number of datasets were used for the automated annotations process and validation. The ability to assess a semi-supervised scenario was unfortunately out of scope due to the many extra variables it would introduce when trying to conduct an unbiased evaluation of this process. The annotation system does already support the ability to manually annotate both attributes and datasets, which has been briefly tested by comparing the process of annotating datasets for both manual and MetaMapper assisted approaches.

The results of this comparison showed slightly faster annotation times (26%). The main cause of this difference can be explained due to the fact that the attribute names had to be typed fully for the manual method. The rest of the process has many similarities, as QGIS also uses GDAL which made the data ingestion process nearly identical. However, apart from annotations times, the MetaMapper assisted approach is still advantageous because it is less error prone. The manual method is fully dependent on the human annotator, and just a spelling mistake could cause data quality issues.

In annotating core concept dataset types, both an aggressive and a conservative strategy may be used. In the aggressive strategy, data types beyond just *ObjectDS* and *LatticeDS* could be predicted by simply picking common relationships between the geometry type and the core concept type. For example, in the gold standard dataset a *LineDataSet* almost exclusively has *NetworkDS* labelled as well. This strategy would greatly increase the recall, at the cost of precision due to a rise in false positives. As is apparent, the conservative strategy was employed instead to reduce the number of assumptions in the annotation system as it has only been tested on the Amsterdam open geodata datasets.

The ability to identify similar attributes and unify them under a single SKOS concept is limited by the distribution comparison for numeric data and the text categorization model for textual data. Both ideas have been briefly mentioned in previous research, but it would still require validation within the context of this research. While the

method seems to work in matching various attributes and leaving out others, its accuracy can not be guaranteed. Strong performance of both methods is required before the annotation system could be regarded as a (simplified) form of active machine learning. Such an approach could open up new possibilities in predicting dataset types outside of just object detection. Previous research (Scheider & Huisjes, 2019) also indicated a machine learning model to predict extensive and intensive measures, which could have tremendous benefits when integrated in the annotation system.

The development work for MetaMapper and the exploration and research into annotating the still relatively novel core concept data types to real world data has proven to be very promising. It may already be used a low bar of entry for new annotators, while future work could improve the automated annotations by improving the identification models in combination with active learning so that all dataset and attribute annotations can be labelled automatically as well.

6 Conclusion

The data processing pipeline for annotating geospatial data from the web can be automated in various levels and degrees. Retrieving and preparing the data is often possible without any human supervision, whereas the annotation process is less suitable for automated processes. Developing an annotation system has helped answer the research questions regarding the subsequent steps from spatial data retrieval to annotation, which in turn has informed the overall research objective in evaluating such a process.

With the creation of MetaMapper, various possibilities to retrieve online spatial datasets were explored and implemented. The first sub-question “*What are the different means of storage for online spatial datasets and in what manner can the source metadata be annotated?*” pertained to annotating the metadata about the dataset publisher and its online location. In an ideal world, datasets would not have to be retrieved by first creating data extraction rules using a wrapper because such information would be standardized across the web. However, in practice each dataset is hosted on different web pages, with different layouts and different means of retrieval. In order to create a unified retrieval method for new datasets, a wrapper allowed both the flexibility of extracting information from any webpage and the maintainability by being able to detect changes to the online publisher. Concluding the first sub-question, retrieving the dataset and its metadata is still difficult due to a lack of standardization for online web pages. Semantic web promises a vision that should deprecate implementations such as the web wrapper in MetaMapper, but until then spatial datasets will still be hosted on non-machine readable web pages that lack the foresight of such a future.

Similar issues exist once data has been pulled from an online spatial resource. The second sub-question “*How can attributes be automatically identified and annotated?*” requires identifying data types and patterns. When dealing with perfectly clean data, such logic would not be difficult to implement. Unfortunately, this is not realistic and data cleaning has grown to become a mandatory step. For example, when dealing with a column that depicts quantities of birds for a given location it makes sense to typecast this attribute as numeric. Detecting this data type is not always as easy as it sounds, because the dataset may contain the word “bird”, which causes it to be categorized as text instead. Such failures are all too common for flawless data type detection.

When the data type of an attribute is known, identifying if it is similar to another group of attributes that are unified under a single SKOS concept is a fuzzy matter. In case of numeric data, distribution comparisons is a possible candidate but does not guarantee a successful match due to unknown nature of the unseen input data. Text categorization using a Naive Bayes classifier will perform better over time as more labelled data corresponds with better predictions. When attribute identification is

able to merge similar columns it will add meaningful links between different datasets that allow inference after just one dataset was annotated manually, or increase the number of data instances to automatically detect core concept data types. With the high probability of unclean data, and the requirement to be able to match any set of attributes, it is desirable to create loose identification logic rather than rigid rules such as minimum and maximum boundaries or regex rules.

Whereas identifying similar attributes is better suited as a fuzzy system, identifying similar geometries benefits from strict enforcement. The third sub-question “*How can geometries be matched to spatial repositories?*” relates to comparing unseen geometries to known places. This is made possible by retrieving large geographic database such as OpenStreetMap. Such a repository exclusively contains real world objects and if a geometry can be matched with an object in the repository it can be inferred that it is also an object. Finding similar geometries has higher accuracy the more vertices the geometry has. Clearly, a point has only one other point to match and compare, which has a higher chance of finding a false positive within the repository. Complex shapes such as administrative boundaries have thousands of vertices, all of which need to be within close proximity to be classified as similar using Hausdorff distance. In both cases, not just a singular geometry but the entire set of geometries for the whole dataset is compared which decreases the possibility for a false positive match on the column as a whole.

The final sub-question “*How can automatic annotations of geospatial concepts be validated?*” requires the entire dataset and its attributes to be annotated with core concept data types. This process is the purpose of the annotation framework and evaluated using a gold standard dataset. Geometry type annotations are the most robust, but also offer the least amount of novel information. Both core concept properties and measurement level annotations are a lot more difficult to automatically label, but when labelled successfully in either a manual or automatic fashion, it adds a valuable semantic dimension that allows matching datasets with the appropriate geocomputational functions. As a baseline, only automatic annotations were evaluated against the gold standard dataset. Because not all core concept data types can be detected from just the data instances this causes either a low precision or a low recall, which is demonstrated by the low score in the harmonized f-measure.

In conclusion, the main research objective “*To what extent can automated processes find, extract, and annotate spatial data from the web with geospatial concepts?*” covers a wide range of topics that can each be solved and automated with various levels of precision. The overarching theme of facilitating geospatial data integration with regards to automated geographic analysis has promising developments, especially regarding a combination of human and machine intelligence. The annotation framework called MetaMapper is able to find and extract geospatial data and lowers the complexity for a human annotator to annotate semantically difficult relations. This should at the very least increase the amount of available annotated geospatial datasets that can be used in future research, and with further development drastic improvements can be made in annotating with ever decreasing human supervision.

7 References

- Atemezing, G. A. (2015). Doctorat ParisTech SE TELECOM ParisTech Publishing and Consuming Geo-Spatial and Government Data on the Semantic Web.
- Berners-Lee, T., Hendler, J. & Lassila, O. (2001). The semantic web. *Scientific American*, 284 (5), 34-43. <https://doi.org/10.1038/scientificamerican0501-34>
- Bizer, C., Eckert, K., Meusel, R., Mühleisen, H., Schuhmacher, M., & Völker, J. (2013). Deployment of RDFa, microdata, and microformats on the web - A quantitative analysis. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 8219 LNCS(PART 2), 17–32. https://doi.org/10.1007/978-3-642-41338-4_2
- Brickley, D., & Guha, R. V. (2014). RDF Schema 1.1. W3C Recommendation, 2014. <http://www.w3.org/TR/2014/REC-rdf-schema-20140225/>
- Chang, C.-H., Kaye, M., Girgis, M. R., & Shaalan, K. F. (2006). A Survey of Web Information Extraction Systems. *IEEE Transactions on Knowledge and Data Engineering*, 18(10), 1411–1428. <https://doi.org/10.1109/TKDE.2006.152>
- Chrisman, N. R. (1998). Rethinking Levels of Measurement for Cartography. *Cartography and Geographic Information Science*, 25(4), 231–242. <https://doi.org/10.1559/152304098782383043>
- Cruz, I. F., Ganesh, V. R., & Mirrezaei, S. I. (2013). Semantic extraction of geographic data from web tables for big data integration. Proceedings of the 7th Workshop on Geographic Information Retrieval, GIR 2013, 19–26. <https://doi.org/10.1145/2533888.2533939>
- Datalift (2019). Available from: <https://datalift.org/> [Accessed November 2019]
- Ferrara, E., De Meo, P., Fiumara, G., & Baumgartner, R. (2014). Web data extraction, applications and techniques: A survey. *Knowledge-Based Systems*, 70, 301–323. <https://doi.org/10.1016/j.knosys.2014.07.007>
- Freelon, D. (2018). Computational Research in the Post-API Age, Political Communication. *Present-Day Corporate Communication*, 35(4), 665–668. <https://doi.org/10.1080/10584609.2018.1477506>
- GeoKnow (2019). Available from: <http://geoknow.eu/Welcome.html> [Accessed November 2019]
- Guarino, N. (1997). Understanding, building and using ontologies. *International Journal of Human- Computer Studies*, 46 (2-3), 293-310. <https://doi.org/10.1006/ijhc.1996.0091>

- Horrocks, I., Patel-Schneider, P. F., & Van Harmelen, F. (2003). From SHIQ and RDF to OWL: The making of a Web Ontology Language. *Web Semantics*, 1(1), 7–26. <https://doi.org/10.1016/j.websem.2003.07.001>
- Jones, C. B., & Purves, R. S. (2008). Geographical information retrieval. *International Journal of Geographical Information Science*, 22(3), 219–228. <https://doi.org/10.1080/13658810701626343>
- Klien, E. (2007). A rule-based strategy for the semantic annotation of geodata. *Transactions in GIS*, 11(3), 437–452. <https://doi.org/10.1111/j.1467-9671.2007.01054.x>
- Knoblock, C. A., Szekely, P., Ambite, J. L., Goel, A., Gupta, S., Lerman, K., ... Mallick, P. (2012). Semi-automatically mapping structured sources into the semantic web. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 7295 LNCS, 375–390. https://doi.org/10.1007/978-3-642-30284-8_32
- Krijnen, D., Bot, R., & Lampropoulos, G. (2004). Automated Web Scraping APIs, 5. Retrieved from http://mediatechnology.leiden.edu/images/uploads/docs/wt2014_web_scraping.pdf
- Kruiger, H., Kasalica, V., Meerlo, R., Lamprecht, A & Scheider, S. (n.d.). Loose programming of GIS workflow with geo-analytical concepts.
- Kuhn, W. (2012). Core concepts of spatial information for transdisciplinary research. *International Journal of Geographical Information Science* 26 (12), 2267–2276. <https://doi.org/10.1080/13658816.2012.722637>
- LinkedGeoData (2019). Available from: <http://linkedgeodata.org/Datasets> [Accessed November 2019]
- Lutz, M., & Klien, E. (2006). Ontology-based retrieval of geographic information. *International Journal of Geographical Information Science*, 20(3), 233–260. <https://doi.org/10.1080/13658810500287107>
- Macário, C. G. N., & Medeiros, C. B. (2009). A framework for semantic annotation of geospatial data for agriculture. *International Journal of Metadata, Semantics and Ontologies*, 4(1–2), 118–132. <https://doi.org/10.1504/IJMSO.2009.026260>
- Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to Information Retrieval*. Cambridge: Cambridge University Press.

- Montello, D. R., Goodchild, M. F., Gottsegen, J., & Fohl, P. (2003). Where's downtown?: Behavioral methods for determining referents of vague spatial queries. *Spatial Cognition and Computation*, 3(2–3), 185–204. <https://doi.org/10.1080/13875868.2003.9683761>
- Neumaier, S., Polleres, A., Published, A., Neumaier, S., & Polleres, A. (2018). ePub WU Institutional Repository Geo-Semantic Labelling of Open Data . SEMANTiCS 2018-14th International Geo-Semantic Labelling of Open Data, (August), 0–12.
- Nowak, J., Nogueras-Iso, J. & Peedell, S. (2005). Issues of multilinguality in creating a European SDI-The perspective for spatial data interoperability. *Proceedings of the 11th EC-GI&GIS Workshop: ESDI: Setting the Framework*, 47-48.
- Presutti, V., & Gangemi, A. (2008). Content Ontology Design Patterns as Practical Building Blocks for Web Ontologies. https://doi.org/10.1007/978-3-540-87877-3_11
- Prudhomme, C., Homburg, T., Ponciano, J. J., Boochs, F., Cruz, C., & Roxin, A. M. (2019). Interpretation and automatic integration of geospatial data into the Semantic Web: Towards a process of automatic geospatial data interpretation, classification and integration using semantic technologies. *Computing*. <https://doi.org/10.1007/s00607-019-00701-y>
- Prudhomme, C., Homburg, T., Ponciano, J. J., Boochs, F., Roxin, A., & Cruz, C. (2017). Automatic integration of spatial data into the semantic web. *WEBIST 2017 - Proceedings of the 13th International Conference on Web Information Systems and Technologies*, 107–115. <https://doi.org/10.5220/0006306601070115>
- Ramnandan, S. K., Mittal, A., Knoblock, C. A., & Szekely, P. (2015). Assigning semantic labels to data sources. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9088, 403–417. https://doi.org/10.1007/978-3-319-18818-8_25
- Reeve, L., & Han, H. (2005). Survey of semantic annotation platforms. *Proceedings of the ACM Symposium on Applied Computing*, 2(August), 1634–1638. <https://doi.org/10.1145/1066677.1067049>
- Scheider, S., & Huisjes, M. D. (2019). Distinguishing extensive and intensive properties for meaningful geocomputation and mapping. *International Journal of Geographical Information Science*, 33(1), 28–54. <https://doi.org/10.1080/13658816.2018.1514120>

- Scheider, S., Meerlo, R., Kasalica, V., & Lamprecht, A. (in-press). Ontology of core concept data types for answering geo-analytical questions.
<http://josis.org/index.php/josis/article/viewArticle/555>
- Tekli, J., Rjeily, A. A., Chbeir, R., Tekli, G., Houngue, P., Yetongnon, K., & Abebe, M. A. (2013). Semantic to intelligent web era, 159–168.
<https://doi.org/10.1145/2536146.2536150>
- Wang, S. & Manning, C. D. (2012). Baselines and Bigrams: Simple, Good Sentiment and Topic Classification. Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics, 90–94.
- Zhang, Y., Chiang, Y. Y., Szekely, P., & Knoblock, C. A. (2013). A semantic approach to retrieving, linking, and integrating heterogeneous geospatial data. ACM International Conference Proceeding Series, 31–37.
<https://doi.org/10.1145/2516911.2516914>

Appendix A

Access URL	Geometry	Entity
https://maps.amsterdam.nl/open_geodata?GELUID_VERKEER	RegionDataSet	ContourDS
https://maps.amsterdam.nl/open_geodata?WONINGWAARDE_2018_INFLATIE	RegionDataSet	ContourDS
https://maps.amsterdam.nl/open_geodata?WELSTAND_NIVEAUS	VectorTessellation	CoverageDS
https://maps.amsterdam.nl/open_geodata?WELSTAND_SYSTEMEN	VectorTessellation	CoverageDS
https://maps.amsterdam.nl/open_geodata?GRONDGEBRUIK_2017	VectorTessellation	CoverageDS
https://maps.amsterdam.nl/open_geodata?OVERNACHTINGSBELEID	VectorTessellation	CoverageDS
https://maps.amsterdam.nl/open_geodata?PARKEERNORMEN	VectorTessellation	CoverageDS
https://maps.amsterdam.nl/open_geodata?WOII	PointDataSet	EventDS
https://maps.amsterdam.nl/open_geodata?STADSLOODS	PointDataSet	EventDS
https://maps.amsterdam.nl/open_geodata?BIJEN	PointDataSet	EventDS
https://maps.amsterdam.nl/open_geodata?EPR_NESTEN	RegionDataSet	EventDS
https://maps.amsterdam.nl/open_geodata/?k=192	VectorTessellation	LatticeDS
https://maps.amsterdam.nl/open_geodata/?k=204	VectorTessellation	LatticeDS
https://maps.amsterdam.nl/open_geodata/?k=154	LineDataSet	LineMeasuresDS
https://maps.amsterdam.nl/open_geodata?VONDELPARK_ZITBANKEN	PointDataSet	ObjectDS
https://maps.amsterdam.nl/open_geodata/?k=273	PointDataSet	ObjectDS
https://maps.amsterdam.nl/open_geodata/?k=232	PointDataSet	ObjectDS
https://maps.amsterdam.nl/open_geodata/?k=87	PointDataSet	ObjectDS
https://maps.amsterdam.nl/open_geodata/?k=93	PointDataSet	ObjectDS
https://maps.amsterdam.nl/open_geodata/?k=62	PointDataSet	ObjectDS
https://maps.amsterdam.nl/open_geodata/?k=228	PointDataSet	ObjectDS
https://maps.amsterdam.nl/open_geodata/?k=96	PointDataSet	ObjectDS
https://maps.amsterdam.nl/open_geodata/?k=51	PointDataSet	ObjectDS
https://maps.amsterdam.nl/open_geodata/?k=136	PointDataSet	ObjectDS
https://maps.amsterdam.nl/open_geodata/?k=58	PointDataSet	ObjectDS
https://maps.amsterdam.nl/open_geodata/?k=123	PointDataSet	ObjectDS
https://maps.amsterdam.nl/open_geodata/?k=68	PointDataSet	ObjectDS
https://maps.amsterdam.nl/open_geodata/?k=155	PointDataSet	ObjectDS
https://maps.amsterdam.nl/open_geodata/?k=141	PointDataSet	ObjectDS
https://maps.amsterdam.nl/open_geodata/?k=272	PointDataSet	ObjectDS
https://maps.amsterdam.nl/open_geodata/?k=240	PointDataSet	ObjectDS

https://maps.amsterdam.nl/open_geodata/?k=264	PointDataSet	ObjectDS
https://maps.amsterdam.nl/open_geodata/?k=251	PointDataSet	ObjectDS
https://maps.amsterdam.nl/open_geodata?SPORT_OPENBAAR	PointDataSet	ObjectDS
https://maps.amsterdam.nl/open_geodata/?k=218	PointDataSet	ObjectDS
https://maps.amsterdam.nl/open_geodata/?k=182	PointDataSet	ObjectDS
https://maps.amsterdam.nl/open_geodata/?k=61	PointDataSet	ObjectDS
https://maps.amsterdam.nl/open_geodata/?k=296	PointDataSet	ObjectDS
https://maps.amsterdam.nl/open_geodata/?k=115	PointDataSet	ObjectDS
https://maps.amsterdam.nl/open_geodata/?k=290	RegionDataSet	ObjectDS
https://maps.amsterdam.nl/open_geodata/?k=279	RegionDataSet	ObjectDS
https://maps.amsterdam.nl/open_geodata/?k=301	RegionDataSet	ObjectDS
https://maps.amsterdam.nl/open_geodata/?k=208	RegionDataSet	ObjectDS
https://maps.amsterdam.nl/open_geodata/?k=166	RegionDataSet	ObjectDS
https://maps.amsterdam.nl/open_geodata/?k=49	RegionDataSet	ObjectDS
https://maps.amsterdam.nl/open_geodata/?k=224	LineDataSet	NetworkDS
https://maps.amsterdam.nl/open_geodata/?k=275	LineDataSet	NetworkDS
https://maps.amsterdam.nl/open_geodata/?k=265	LineDataSet	NetworkDS
https://maps.amsterdam.nl/open_geodata/?k=233	LineDataSet	NetworkDS
https://maps.amsterdam.nl/open_geodata/?k=84	LineDataSet	NetworkDS
https://maps.amsterdam.nl/open_geodata/?k=220	LineDataSet	NetworkDS
https://maps.amsterdam.nl/open_geodata?WALKABILITY	LineDataSet	NetworkDS
https://maps.amsterdam.nl/open_geodata/?k=63	PointDataSet	ObjectDS
https://maps.amsterdam.nl/open_geodata/?k=99	RegionDataSet	ObjectDS
https://maps.amsterdam.nl/open_geodata/?k=106	PointDataSet	PointMeasuresDS
https://maps.amsterdam.nl/open_geodata/?k=319	RegionDataSet	PatchDS
https://maps.amsterdam.nl/open_geodata/?k=57	VectorTessellation	PatchDS
https://maps.amsterdam.nl/open_geodata?HOOFDGROENSTRUCTUUR	RegionDataSet	ObjectDS
https://maps.amsterdam.nl/open_geodata/?k=242	RegionDataSet	ObjectDS
https://maps.amsterdam.nl/open_geodata/?k=247	PointDataSet	ObjectDS

Appendix B

access_url	column_name	attribute_type
https://maps.amsterdam.nl/open_geodata/?k=265	Besluitnummer	NominalA
https://maps.amsterdam.nl/open_geodata/?k=265	Omschrijving	
https://maps.amsterdam.nl/open_geodata/?k=265	Pijlen	
https://maps.amsterdam.nl/open_geodata/?k=265	SELECTIE	NominalA
https://maps.amsterdam.nl/open_geodata/?k=265	Status	NominalA
https://maps.amsterdam.nl/open_geodata?WOII	Adres	
https://maps.amsterdam.nl/open_geodata?WOII	Datum	
https://maps.amsterdam.nl/open_geodata?WOII	Foto	
https://maps.amsterdam.nl/open_geodata?WOII	Omschrijving	
https://maps.amsterdam.nl/open_geodata?WOII	Soort	NominalA
https://maps.amsterdam.nl/open_geodata/?k=233	AUTO	NominalA
https://maps.amsterdam.nl/open_geodata/?k=233	BREED_VERVOER	
https://maps.amsterdam.nl/open_geodata/?k=233	BUS	NominalA
https://maps.amsterdam.nl/open_geodata/?k=233	FIETS	NominalA
https://maps.amsterdam.nl/open_geodata/?k=233	GME_NAAM	
https://maps.amsterdam.nl/open_geodata/?k=233	LINKNR	
https://maps.amsterdam.nl/open_geodata/?k=233	MAX_SNELHEID	OrdinalA
https://maps.amsterdam.nl/open_geodata/?k=233	RIJRICHTINGEN	NominalA
https://maps.amsterdam.nl/open_geodata/?k=233	STT_NAAM	
https://maps.amsterdam.nl/open_geodata/?k=233	TRAM	NominalA
https://maps.amsterdam.nl/open_geodata/?k=233	VOET	NominalA
https://maps.amsterdam.nl/open_geodata/?k=233	WPSNAAMNEN	
https://maps.amsterdam.nl/open_geodata/?k=233	WVK_ID	
https://maps.amsterdam.nl/open_geodata/?k=296	BOOM_ID	
https://maps.amsterdam.nl/open_geodata/?k=296	BOOMSOORT	
https://maps.amsterdam.nl/open_geodata/?k=296	DIAMETERKLASSE	
https://maps.amsterdam.nl/open_geodata/?k=296	HOOGTEKLASSE	
https://maps.amsterdam.nl/open_geodata/?k=296	PLANTJAAR	
https://maps.amsterdam.nl/open_geodata/?k=136	Aantal_kasten	
https://maps.amsterdam.nl/open_geodata/?k=136	Aantal_volken	RatioA
https://maps.amsterdam.nl/open_geodata?WONINGWAARDE 2018 INFLATIE	Lower	RatioA
https://maps.amsterdam.nl/open_geodata?WONINGWAARDE 2018 INFLATIE	Upper	RatioA
https://maps.amsterdam.nl/open_geodata?SPORT OPENBAAR	Foto	
https://maps.amsterdam.nl/open_geodata?SPORT OPENBAAR	Locatie	
https://maps.amsterdam.nl/open_geodata?SPORT OPENBAAR	Naam	

https://maps.amsterdam.nl/open_geodata?SPORT_OPENBAAR	Omheining	NominalA
https://maps.amsterdam.nl/open_geodata?SPORT_OPENBAAR	Omschrijving	
https://maps.amsterdam.nl/open_geodata?SPORT_OPENBAAR	Ondergrond	NominalA
https://maps.amsterdam.nl/open_geodata?SPORT_OPENBAAR	Oppervlak	RatioA
https://maps.amsterdam.nl/open_geodata?SPORT_OPENBAAR	Overdekt	BooleanA
https://maps.amsterdam.nl/open_geodata?SPORT_OPENBAAR	Sportvoorziening	NominalA
https://maps.amsterdam.nl/open_geodata?SPORT_OPENBAAR	Verlichting	BooleanA
https://maps.amsterdam.nl/open_geodata/?k=208	Bouwjaar	OrdinalA
https://maps.amsterdam.nl/open_geodata/?k=208	CORP	NominalA
https://maps.amsterdam.nl/open_geodata/?k=208	Corporatie	NominalA
https://maps.amsterdam.nl/open_geodata/?k=208	Corporatie_woningen	RatioA
https://maps.amsterdam.nl/open_geodata/?k=208	Onzelfstandig	RatioA
https://maps.amsterdam.nl/open_geodata/?k=208	PERC	IntervalA
https://maps.amsterdam.nl/open_geodata/?k=208	Percentage	IntervalA
https://maps.amsterdam.nl/open_geodata/?k=208	Woningen_in_blok	RatioA
https://maps.amsterdam.nl/open_geodata/?k=208	Zelfstandig	RatioA
https://maps.amsterdam.nl/open_geodata/?k=123	Artikelen	NominalA
https://maps.amsterdam.nl/open_geodata/?k=123	Dagen	
https://maps.amsterdam.nl/open_geodata/?k=123	Dinsdag	
https://maps.amsterdam.nl/open_geodata/?k=123	Donderdag	
https://maps.amsterdam.nl/open_geodata/?k=123	Gemeente	BooleanA
https://maps.amsterdam.nl/open_geodata/?k=123	Locatie	
https://maps.amsterdam.nl/open_geodata/?k=123	Maandag	
https://maps.amsterdam.nl/open_geodata/?k=123	Meer_informatie	
https://maps.amsterdam.nl/open_geodata/?k=123	Vrijdag	
https://maps.amsterdam.nl/open_geodata/?k=123	Website	
https://maps.amsterdam.nl/open_geodata/?k=123	Woensdag	
https://maps.amsterdam.nl/open_geodata/?k=123	Zaterdag	
https://maps.amsterdam.nl/open_geodata/?k=123	Zondag	
https://maps.amsterdam.nl/open_geodata/?k=319	BBQ	NominalA
https://maps.amsterdam.nl/open_geodata/?k=319	Naam	
https://maps.amsterdam.nl/open_geodata/?k=319	Toelichting_bbq	
https://maps.amsterdam.nl/open_geodata/?k=106	Aantal_in_2009	RatioA
https://maps.amsterdam.nl/open_geodata/?k=106	Aantal_in_2010	RatioA
https://maps.amsterdam.nl/open_geodata/?k=106	Aantal_in_2011	RatioA
https://maps.amsterdam.nl/open_geodata/?k=106	Aantal_in_2012	RatioA
https://maps.amsterdam.nl/open_geodata/?k=106	Aantal_in_2013	RatioA
https://maps.amsterdam.nl/open_geodata/?k=106	Beschermd	NominalA
https://maps.amsterdam.nl/open_geodata/?k=106	Brugnummer	

https://maps.amsterdam.nl/open_geodata/?k=106	Kademuur	BooleanA
https://maps.amsterdam.nl/open_geodata/?k=106	Nederlandse_naam	
https://maps.amsterdam.nl/open_geodata/?k=106	Rode_lijst	
https://maps.amsterdam.nl/open_geodata/?k=106	SELECTIE	
https://maps.amsterdam.nl/open_geodata/?k=106	Straat	
https://maps.amsterdam.nl/open_geodata/?k=106	Water	
https://maps.amsterdam.nl/open_geodata/?k=106	Wetenschappelijke_naam	
https://maps.amsterdam.nl/open_geodata/?k=87	Aantal_2011	RatioA
https://maps.amsterdam.nl/open_geodata/?k=87	Aantal_2012	RatioA
https://maps.amsterdam.nl/open_geodata/?k=87	Aantal_2013	RatioA
https://maps.amsterdam.nl/open_geodata/?k=87	Aantal_2014	RatioA
https://maps.amsterdam.nl/open_geodata/?k=87	Aantal_2015	RatioA
https://maps.amsterdam.nl/open_geodata/?k=87	Aantal_2016	RatioA
https://maps.amsterdam.nl/open_geodata/?k=87	Aantal_2017	RatioA
https://maps.amsterdam.nl/open_geodata/?k=87	Aantal_2018	RatioA
https://maps.amsterdam.nl/open_geodata/?k=87	Aantal_2019	RatioA
https://maps.amsterdam.nl/open_geodata/?k=87	Foto1	
https://maps.amsterdam.nl/open_geodata/?k=87	Foto2	
https://maps.amsterdam.nl/open_geodata/?k=87	Huisnummer	
https://maps.amsterdam.nl/open_geodata/?k=87	Straatnaam	
https://maps.amsterdam.nl/open_geodata/?k=87	Verblijfplaats	
https://maps.amsterdam.nl/open_geodata/?k=87	Vogel	NominalA
https://maps.amsterdam.nl/open_geodata/?k=87	Voorziening	
https://maps.amsterdam.nl/open_geodata/?k=272	Adres_gebouw	
https://maps.amsterdam.nl/open_geodata/?k=272	Afbeelding	
https://maps.amsterdam.nl/open_geodata/?k=272	Architectuurhistorische_waarde	
https://maps.amsterdam.nl/open_geodata/?k=272	Bouwjaar	OrdinalA
https://maps.amsterdam.nl/open_geodata/?k=272	Cultuurhistorische_waarde	
https://maps.amsterdam.nl/open_geodata/?k=272	Jaartal	OrdinalA
https://maps.amsterdam.nl/open_geodata/?k=272	Kunstenaar	
https://maps.amsterdam.nl/open_geodata/?k=272	Kunsthistorische_waarde	
https://maps.amsterdam.nl/open_geodata/?k=272	Status_gebouw	
https://maps.amsterdam.nl/open_geodata/?k=272	Techniek	NominalA
https://maps.amsterdam.nl/open_geodata/?k=272	Titel_kunstwerk	
https://maps.amsterdam.nl/open_geodata/?k=272	Waardering	OrdinalA
https://maps.amsterdam.nl/open_geodata/?k=51	Adres	
https://maps.amsterdam.nl/open_geodata/?k=51	Foto	
https://maps.amsterdam.nl/open_geodata/?k=51	Oppervlakte_m2	RatioA
https://maps.amsterdam.nl/open_geodata/?k=51	SELECTIE	NominalA
https://maps.amsterdam.nl/open_geodata/?k=51	Sinds	OrdinalA
https://maps.amsterdam.nl/open_geodata/?k=51	Soort_tuin	NominalA
https://maps.amsterdam.nl/open_geodata/?k=51	Stadsdeel	
https://maps.amsterdam.nl/open_geodata/?k=51	Volgnummer	
https://maps.amsterdam.nl/open_geodata/?k=51	Weblink	
https://maps.amsterdam.nl/open_geodata/?k=275	Lijnnummer	

https://maps.amsterdam.nl/open_geodata/?k=275	Type_hoofdstraat	
https://maps.amsterdam.nl/open_geodata/?k=275	Type_stadsstraat	
https://maps.amsterdam.nl/open_geodata/?k=63	Beheergebied	
https://maps.amsterdam.nl/open_geodata/?k=63	Beschrijving_knelpunt	
https://maps.amsterdam.nl/open_geodata/?k=63	Beschrijving_oplossing	
https://maps.amsterdam.nl/open_geodata/?k=63	Detail	
https://maps.amsterdam.nl/open_geodata/?k=63	Diersoorten	NominalA
https://maps.amsterdam.nl/open_geodata/?k=63	FILTER	NominalA
https://maps.amsterdam.nl/open_geodata/?k=63	Foto	
https://maps.amsterdam.nl/open_geodata/?k=63	Geschatte_kosten	RatioA
https://maps.amsterdam.nl/open_geodata/?k=63	Jaar_realisatie	OrdinalA
https://maps.amsterdam.nl/open_geodata/?k=63	Knelpuntnummer	
https://maps.amsterdam.nl/open_geodata/?k=63	Locatie	
https://maps.amsterdam.nl/open_geodata/?k=63	SELECTIE	NominalA
https://maps.amsterdam.nl/open_geodata/?k=63	Status	NominalA
https://maps.amsterdam.nl/open_geodata/?k=63	Type_knelpunt	NominalA
https://maps.amsterdam.nl/open_geodata/?k=63	Type_oplossing	NominalA
https://maps.amsterdam.nl/open_geodata/?k=154	COMBI	
https://maps.amsterdam.nl/open_geodata/?k=154	Gebied	
https://maps.amsterdam.nl/open_geodata/?k=154	Gebied_voo	
https://maps.amsterdam.nl/open_geodata/?k=154	Iets_voorg	
https://maps.amsterdam.nl/open_geodata/?k=154	Objectnumm	
https://maps.amsterdam.nl/open_geodata/?k=154	Opmerking	
https://maps.amsterdam.nl/open_geodata/?k=154	Redactie	
https://maps.amsterdam.nl/open_geodata/?k=154	SELECTIE	NominalA
https://maps.amsterdam.nl/open_geodata/?k=154	SELECTIE_G	NominalA
https://maps.amsterdam.nl/open_geodata/?k=154	Stadsdeel	
https://maps.amsterdam.nl/open_geodata/?k=154	Zone	
https://maps.amsterdam.nl/open_geodata/?k=154	Zone_voors	
https://maps.amsterdam.nl/open_geodata/?k=228	Aantal_bewoners	RatioA
https://maps.amsterdam.nl/open_geodata/?k=228	Email	
https://maps.amsterdam.nl/open_geodata/?k=228	FILTER	NominalA
https://maps.amsterdam.nl/open_geodata/?k=228	Foto	
https://maps.amsterdam.nl/open_geodata/?k=228	Gebruik	NominalA
https://maps.amsterdam.nl/open_geodata/?k=228	Initiatiefnemer	
https://maps.amsterdam.nl/open_geodata/?k=228	Soort_afval	NominalA
https://maps.amsterdam.nl/open_geodata/?k=228	Startjaar	OrdinalA
https://maps.amsterdam.nl/open_geodata/?k=228	Straatnaam	
https://maps.amsterdam.nl/open_geodata/?k=301	NETTYPE	NominalA
https://maps.amsterdam.nl/open_geodata/?k=301	STATUS	NominalA
https://maps.amsterdam.nl/open_geodata/?k=84	plein_park	BooleanA
https://maps.amsterdam.nl/open_geodata/?k=141	Regel	NominalA
https://maps.amsterdam.nl/open_geodata/?k=141	SELECTIE	NominalA
https://maps.amsterdam.nl/open_geodata/?k=240	Dagen_geopend	
https://maps.amsterdam.nl/open_geodata/?k=240	Foto	
https://maps.amsterdam.nl/open_geodata/?k=240	Omschrijving	

https://maps.amsterdam.nl/open_geodata/?k=240	Openingstijden	
https://maps.amsterdam.nl/open_geodata/?k=240	Prijs_per_gebruik	RatioA
https://maps.amsterdam.nl/open_geodata/?k=240	SELECTIE	NominalA
https://maps.amsterdam.nl/open_geodata/?k=240	Soort	
https://maps.amsterdam.nl/open_geodata?WELSTAND_NIVEAUS	Niveau	NominalA
https://maps.amsterdam.nl/open_geodata?WELSTAND_NIVEAUS	SELECTIE	NominalA
https://maps.amsterdam.nl/open_geodata/?k=242	Categorie	NominalA
https://maps.amsterdam.nl/open_geodata?PARKEERNORMEN	Categorie	NominalA
https://maps.amsterdam.nl/open_geodata/?k=68	Aanwijzingsdatum	
https://maps.amsterdam.nl/open_geodata/?k=68	Monumentenregister	
https://maps.amsterdam.nl/open_geodata/?k=68	Monumentnummer	
https://maps.amsterdam.nl/open_geodata/?k=68	Naam	
https://maps.amsterdam.nl/open_geodata/?k=68	Status	NominalA
https://maps.amsterdam.nl/open_geodata/?k=68	Weblink	
https://maps.amsterdam.nl/open_geodata?GEBIED_BUURTEN	Buurt	
https://maps.amsterdam.nl/open_geodata?GEBIED_BUURTEN	Buurt_code	
https://maps.amsterdam.nl/open_geodata?GEBIED_BUURTEN	Buurtcombinatie_code	
https://maps.amsterdam.nl/open_geodata?GEBIED_BUURTEN	Opp_m2	RatioA
https://maps.amsterdam.nl/open_geodata?GEBIED_BUURTEN	Stadsdeel_code	
https://maps.amsterdam.nl/open_geodata/?k=264	Actief	BooleanA
https://maps.amsterdam.nl/open_geodata/?k=264	Objectnummer	
https://maps.amsterdam.nl/open_geodata/?k=264	Privacyverklaring	
https://maps.amsterdam.nl/open_geodata/?k=264	Rotatie	
https://maps.amsterdam.nl/open_geodata/?k=264	Soort	NominalA
https://maps.amsterdam.nl/open_geodata/?k=264	Voeding	NominalA
https://maps.amsterdam.nl/open_geodata/?k=232	Boomnummer	
https://maps.amsterdam.nl/open_geodata/?k=232	Boomsoort	NominalA
https://maps.amsterdam.nl/open_geodata/?k=232	Herplantsoort	NominalA
https://maps.amsterdam.nl/open_geodata/?k=232	Maatregel	NominalA
https://maps.amsterdam.nl/open_geodata/?k=232	Maatregel_extra	NominalA
https://maps.amsterdam.nl/open_geodata/?k=232	Planning_herplant	
https://maps.amsterdam.nl/open_geodata/?k=232	Planning_kap_snoei	
https://maps.amsterdam.nl/open_geodata/?k=232	Stadsdeel	
https://maps.amsterdam.nl/open_geodata/?k=232	Takbreuk	
https://maps.amsterdam.nl/open_geodata?GRONDGEBRUIK_2017	AMScore	NominalA
https://maps.amsterdam.nl/open_geodata?GRONDGEBRUIK_2017	AMScore_Oms	NominalA

https://maps.amsterdam.nl/open_geodata?GRONDGEBRUIK_2017	AMScore_OmsEng	NominalA
https://maps.amsterdam.nl/open_geodata?GRONDGEBRUIK_2017	CBScode1	NominalA
https://maps.amsterdam.nl/open_geodata?GRONDGEBRUIK_2017	CBScode1_Oms	NominalA
https://maps.amsterdam.nl/open_geodata?GRONDGEBRUIK_2017	CBScode1_OmsEng	NominalA
https://maps.amsterdam.nl/open_geodata?GRONDGEBRUIK_2017	CBScode2	NominalA
https://maps.amsterdam.nl/open_geodata?GRONDGEBRUIK_2017	CBScode2_Oms	NominalA
https://maps.amsterdam.nl/open_geodata?GRONDGEBRUIK_2017	CBScode2_OmsEng	NominalA
https://maps.amsterdam.nl/open_geodata?GRONDGEBRUIK_2017	Oppervlakte_m2	RatioA
https://maps.amsterdam.nl/open_geodata/?k=251	Functie	BooleanA
https://maps.amsterdam.nl/open_geodata/?k=251	Gedetecteerde_panellen	RatioA
https://maps.amsterdam.nl/open_geodata/?k=251	Vermogen	RatioA
https://maps.amsterdam.nl/open_geodata/?k=273	Adres	
https://maps.amsterdam.nl/open_geodata/?k=273	Bijlschrift	
https://maps.amsterdam.nl/open_geodata/?k=273	Foto_object	
https://maps.amsterdam.nl/open_geodata/?k=273	Foto_put	
https://maps.amsterdam.nl/open_geodata/?k=273	Legenda	NominalA
https://maps.amsterdam.nl/open_geodata/?k=273	SELECTIE	NominalA
https://maps.amsterdam.nl/open_geodata/?k=273	Website	
https://maps.amsterdam.nl/open_geodata/?k=96	Locatie	NominalA
https://maps.amsterdam.nl/open_geodata/?k=96	Vermogen_MW	RatioA
https://maps.amsterdam.nl/open_geodata/?k=96	Verwachte_Opbrengst_GW h_Jaar	RatioA
https://maps.amsterdam.nl/open_geodata/?k=99	Naam	
https://maps.amsterdam.nl/open_geodata/?k=99	Oppervlakte_m2	RatioA
https://maps.amsterdam.nl/open_geodata/?k=99	Stadsdeel	NominalA
https://maps.amsterdam.nl/open_geodata/?k=99	Stadspark	BooleanA
https://maps.amsterdam.nl/open_geodata?EPR_NESTEN	Jaar	OrdinalA
https://maps.amsterdam.nl/open_geodata?EPR_NESTEN	Objectnummer	
https://maps.amsterdam.nl/open_geodata?EPR_NESTEN	Plaagdruk	BooleanA
https://maps.amsterdam.nl/open_geodata?STADSLOODS	Activiteit	
https://maps.amsterdam.nl/open_geodata?STADSLOODS	Initiatiefnemer	
https://maps.amsterdam.nl/open_geodata?STADSLOODS	Jaar	OrdinalA

https://maps.amsterdam.nl/open_geodata?STADSLOODS	Sector	NominalA
https://maps.amsterdam.nl/open_geodata/?k=204	Gebied	
https://maps.amsterdam.nl/open_geodata/?k=204	Gebied_code	
https://maps.amsterdam.nl/open_geodata/?k=204	Opp_m2	RatioA
https://maps.amsterdam.nl/open_geodata/?k=204	Stadsdeel_code	
https://maps.amsterdam.nl/open_geodata/?k=218	Categorie	NominalA
https://maps.amsterdam.nl/open_geodata/?k=218	Naam_locatie	
https://maps.amsterdam.nl/open_geodata?WALKABILITY	Drukte_algemeen	OrdinalA
https://maps.amsterdam.nl/open_geodata?WALKABILITY	Drukte_recreatief	OrdinalA
https://maps.amsterdam.nl/open_geodata?WALKABILITY	Drukte_toeristisch	OrdinalA
https://maps.amsterdam.nl/open_geodata?WALKABILITY	Drukte_werkgerelateerd	OrdinalA
https://maps.amsterdam.nl/open_geodata?WALKABILITY	Loopruimte	OrdinalA
https://maps.amsterdam.nl/open_geodata?WALKABILITY	Straatvaknummer	
https://maps.amsterdam.nl/open_geodata?WALKABILITY	Walkability_algemeen	OrdinalA
https://maps.amsterdam.nl/open_geodata?WALKABILITY	Walkability_recreatief	OrdinalA
https://maps.amsterdam.nl/open_geodata?WALKABILITY	Walkability_toeristisch	OrdinalA
https://maps.amsterdam.nl/open_geodata?WALKABILITY	Walkability_werkgerelateerd	OrdinalA
https://maps.amsterdam.nl/open_geodata?GELUID VERKEER	KLASSE	OrdinalA
https://maps.amsterdam.nl/open_geodata?GELUID VERKEER	SELECTIE	OrdinalA
https://maps.amsterdam.nl/open_geodata?GELUID VERKEER	SOORT	NominalA
https://maps.amsterdam.nl/open_geodata/?k=182	Bouwjaar	OrdinalA
https://maps.amsterdam.nl/open_geodata/?k=182	Objectnummer_Amsterdam	
https://maps.amsterdam.nl/open_geodata/?k=182	Objectnummer_leverancier	
https://maps.amsterdam.nl/open_geodata/?k=182	Rotatie	
https://maps.amsterdam.nl/open_geodata/?k=182	Soort	NominalA
https://maps.amsterdam.nl/open_geodata/?k=182	Soortcode	NominalA
https://maps.amsterdam.nl/open_geodata/?k=182	Standplaats	
https://maps.amsterdam.nl/open_geodata/?k=182	Voeding	NominalA
https://maps.amsterdam.nl/open_geodata/?k=182	VRI_nummer	
https://maps.amsterdam.nl/open_geodata/?k=224	NETTYPE	NominalA
https://maps.amsterdam.nl/open_geodata/?k=224	STATUS	NominalA
https://maps.amsterdam.nl/open_geodata?BIJEN	Gebied	

https://maps.amsterdam.nl/open_geodata?BIJEN	Jaar	OrdinalA
https://maps.amsterdam.nl/open_geodata?BIJEN	Waargenomen_bijen	NominalA
https://maps.amsterdam.nl/open_geodata?VONDELPARK_ZITBANKEN	Banknummer	
https://maps.amsterdam.nl/open_geodata?VONDELPARK_ZITBANKEN	Categorie	NominalA
https://maps.amsterdam.nl/open_geodata?VONDELPARK_ZITBANKEN	Definitief_foto	
https://maps.amsterdam.nl/open_geodata?VONDELPARK_ZITBANKEN	Definitief_type	NominalA
https://maps.amsterdam.nl/open_geodata?VONDELPARK_ZITBANKEN	Geadopteerd	NominalA
https://maps.amsterdam.nl/open_geodata?VONDELPARK_ZITBANKEN	Huidig_type	NominalA
https://maps.amsterdam.nl/open_geodata?VONDELPARK_ZITBANKEN	Rotatie	
https://maps.amsterdam.nl/open_geodata?VONDELPARK_ZITBANKEN	Tekstplaat	
https://maps.amsterdam.nl/open_geodata/?k=290	Drainage	NominalA
https://maps.amsterdam.nl/open_geodata/?k=290	Locatie	
https://maps.amsterdam.nl/open_geodata/?k=61	Foto	
https://maps.amsterdam.nl/open_geodata/?k=61	Geregeld_sinds_schatting	OrdinalA
https://maps.amsterdam.nl/open_geodata/?k=61	Koppeling	NominalA
https://maps.amsterdam.nl/open_geodata/?k=61	Kruispunt	
https://maps.amsterdam.nl/open_geodata/?k=61	Kruispuntnummer	
https://maps.amsterdam.nl/open_geodata/?k=61	Regeling	NominalA
https://maps.amsterdam.nl/open_geodata/?k=61	Soort	NominalA
https://maps.amsterdam.nl/open_geodata/?k=61	Status	BooleanA
https://maps.amsterdam.nl/open_geodata/?k=61	Vervallen_sinds_schatting	OrdinalA
https://maps.amsterdam.nl/open_geodata/?k=61	Weblink	
https://maps.amsterdam.nl/open_geodata/?k=115	Aantal_woningen	RatioA
https://maps.amsterdam.nl/open_geodata/?k=115	Doelgroep	NominalA
https://maps.amsterdam.nl/open_geodata/?k=115	Eigenaar_beheerder	
https://maps.amsterdam.nl/open_geodata/?k=115	Informatie	
https://maps.amsterdam.nl/open_geodata/?k=115	Naam	
https://maps.amsterdam.nl/open_geodata/?k=115	Status	BooleanA
https://maps.amsterdam.nl/open_geodata/?k=115	Straatnaam	
https://maps.amsterdam.nl/open_geodata/?k=115	Termijn_project	NominalA
https://maps.amsterdam.nl/open_geodata/?k=115	Webpagina	
https://maps.amsterdam.nl/open_geodata/?k=49	Aantal_adressen	
https://maps.amsterdam.nl/open_geodata/?k=49	Adres	
https://maps.amsterdam.nl/open_geodata/?k=49	Footprint	
https://maps.amsterdam.nl/open_geodata/?k=49	Functie	NominalA
https://maps.amsterdam.nl/open_geodata/?k=49	Functie_gecorrigeerd	NominalA
https://maps.amsterdam.nl/open_geodata/?k=49	Gebruiksdoel	NominalA
https://maps.amsterdam.nl/open_geodata/?k=49	Gebruiksoppervlakte	RatioA

https://maps.amsterdam.nl/open_geodata/?k=49	Hoofd functie	NominalA
https://maps.amsterdam.nl/open_geodata/?k=49	Hoofd functie_ID	NominalA
https://maps.amsterdam.nl/open_geodata/?k=49	Oppervlakte_gecorrigeerd	
https://maps.amsterdam.nl/open_geodata/?k=49	VOT_ID	
https://maps.amsterdam.nl/open_geodata/?k=49	WOZ_gebruik	NominalA
https://maps.amsterdam.nl/open_geodata?HOOFDGROENSTRUCTUUR	Groentype	NominalA
https://maps.amsterdam.nl/open_geodata?HOOFDGROENSTRUCTUUR	Naam	
https://maps.amsterdam.nl/open_geodata?HOOFDGROENSTRUCTUUR	Nummer	
https://maps.amsterdam.nl/open_geodata?HOOFDGROENSTRUCTUUR	Oppervlakte_ha	RatioA
https://maps.amsterdam.nl/open_geodata?HOOFDGROENSTRUCTUUR	Oppervlakte_m2	RatioA
https://maps.amsterdam.nl/open_geodata?HOOFDGROENSTRUCTUUR	PDF	
https://maps.amsterdam.nl/open_geodata?HOOFDGROENSTRUCTUUR	SELECTIE	NominalA
https://maps.amsterdam.nl/open_geodata/?k=279	Informatie_pdf	
https://maps.amsterdam.nl/open_geodata/?k=279	Ontwikkelbuurtcode	
https://maps.amsterdam.nl/open_geodata/?k=279	Ontwikkelbuurtnaam	
https://maps.amsterdam.nl/open_geodata/?k=279	Stadsdeel	NominalA
https://maps.amsterdam.nl/open_geodata/?k=279	Webpagina	
https://maps.amsterdam.nl/open_geodata/?k=58	Adres	
https://maps.amsterdam.nl/open_geodata/?k=58	BOERBRON	BooleanA
https://maps.amsterdam.nl/open_geodata/?k=58	BOERENMARKT	BooleanA
https://maps.amsterdam.nl/open_geodata/?k=58	Dagen	
https://maps.amsterdam.nl/open_geodata/?k=58	HORECA	BooleanA
https://maps.amsterdam.nl/open_geodata/?k=58	Horeca_soort	
https://maps.amsterdam.nl/open_geodata/?k=58	NATUURWINKEL	BooleanA
https://maps.amsterdam.nl/open_geodata/?k=58	Producten	
https://maps.amsterdam.nl/open_geodata/?k=58	Tijden	
https://maps.amsterdam.nl/open_geodata/?k=58	Website	
https://maps.amsterdam.nl/open_geodata/?k=58	Zaaknaam	
https://maps.amsterdam.nl/open_geodata/?k=57	Bodemfunctie	NominalA
https://maps.amsterdam.nl/open_geodata/?k=57	Diepe_laag	NominalA
https://maps.amsterdam.nl/open_geodata/?k=57	Gebiednummer	
https://maps.amsterdam.nl/open_geodata/?k=57	Locatie	
https://maps.amsterdam.nl/open_geodata/?k=57	Oorspronkelijk_maaiveld	NominalA
https://maps.amsterdam.nl/open_geodata/?k=57	Statistische_kentallen	
https://maps.amsterdam.nl/open_geodata/?k=57	Statistische_kentallen_weg	
https://maps.amsterdam.nl/open_geodata/?k=57	Toelichting	
https://maps.amsterdam.nl/open_geodata/?k=57	Toelichting_weg	
https://maps.amsterdam.nl/open_geodata/?k=57	Toplaag	NominalA
https://maps.amsterdam.nl/open_geodata/?k=57	Zone	NominalA

https://maps.amsterdam.nl/open_geodata/?k=57	Zone_weg	NominalA
https://maps.amsterdam.nl/open_geodata/?k=220	Lijn	NominalA
https://maps.amsterdam.nl/open_geodata/?k=220	Lijn_select	NominalA
https://maps.amsterdam.nl/open_geodata/?k=220	Modaliteit	NominalA
https://maps.amsterdam.nl/open_geodata/?k=192	Opp_m2	RatioA
https://maps.amsterdam.nl/open_geodata/?k=192	Postcode4	NominalA
https://maps.amsterdam.nl/open_geodata/?k=93	Aantasting	NominalA
https://maps.amsterdam.nl/open_geodata/?k=93	Kastanje soort	NominalA
https://maps.amsterdam.nl/open_geodata/?k=93	Nummer	
https://maps.amsterdam.nl/open_geodata?OVERNACHTINGSBELEID	Buurtnaam	
https://maps.amsterdam.nl/open_geodata?OVERNACHTINGSBELEID	SELECTIE	NominalA
https://maps.amsterdam.nl/open_geodata?OVERNACHTINGSBELEID	Voorwaarden	
https://maps.amsterdam.nl/open_geodata/?k=62	Adres	
https://maps.amsterdam.nl/open_geodata/?k=62	Bepantingsituatie	NominalA
https://maps.amsterdam.nl/open_geodata/?k=62	Initiatiefnemer	
https://maps.amsterdam.nl/open_geodata/?k=62	Naam	
https://maps.amsterdam.nl/open_geodata/?k=62	Soort	NominalA
https://maps.amsterdam.nl/open_geodata/?k=62	Status	BooleanA
https://maps.amsterdam.nl/open_geodata/?k=62	Website	
https://maps.amsterdam.nl/open_geodata/?k=166	GEMEENTE_ID	
https://maps.amsterdam.nl/open_geodata/?k=155	AanwijzingsDatum	
https://maps.amsterdam.nl/open_geodata/?k=155	Adres	
https://maps.amsterdam.nl/open_geodata/?k=155	Adressen	
https://maps.amsterdam.nl/open_geodata/?k=155	Afbeelding	
https://maps.amsterdam.nl/open_geodata/?k=155	Monumentnummer	
https://maps.amsterdam.nl/open_geodata/?k=155	Naam	
https://maps.amsterdam.nl/open_geodata/?k=155	Omschrijving	
https://maps.amsterdam.nl/open_geodata/?k=155	SELECTIE	NominalA
https://maps.amsterdam.nl/open_geodata/?k=155	Soort	NominalA
https://maps.amsterdam.nl/open_geodata/?k=155	Status	
https://maps.amsterdam.nl/open_geodata/?k=155	Unescogebed	BooleanA
https://maps.amsterdam.nl/open_geodata/?k=247	OG_FOTO_01	
https://maps.amsterdam.nl/open_geodata/?k=247	OG_FOTO_02	
https://maps.amsterdam.nl/open_geodata/?k=247	OG_FOTO_SC	
https://maps.amsterdam.nl/open_geodata/?k=247	OS_OPPERVL	RatioA
https://maps.amsterdam.nl/open_geodata/?k=247	OS_SOORTNA	NominalA
https://maps.amsterdam.nl/open_geodata?WELSTAND_SYSTEMEN	SELECTIE	NominalA
https://maps.amsterdam.nl/open_geodata?WELSTAND_SYSTEMEN	Systeemnaam	NominalA