

Stairwalker

User Manual

Document authors:

- DENNIS MULLER (UT)
- JOCHEM ELSINGA (UT)
- MAURICE VAN KEULEN (UT)

Software developers:

- ANDREAS WOMBACHER (UT)
- JAN FLOKSTRA (UT)
- HENKE PONS (ARCADIS)
- NIEK DE VRIES (ARCADIS)

The logo consists of the word "COMMIT/" in a bold, black, sans-serif font. The text is centered within a thin black rectangular border.

21 December 2015

UNIVERSITY OF TWENTE.

Contents

1	Introduction	3
2	Set Up	5
2.1	Database Setup	5
2.1.1	PostgreSQL Database Manager Setup	5
2.1.2	PostgreSQL Configuration	6
2.1.3	PostGIS Configuration in PostgreSQL	6
2.1.4	Serverside Stairwalker Extension in PostgreSQL	7
2.2	Pre-Aggregate Database Table	7
2.2.1	Description of Process	7
2.2.2	PreAggregate Tool	8
2.2.3	Current Status Support of Axis Types	9
2.2.4	Current Status Support of Aggregation Types	9
2.3	Tomcat Server Setup	10
2.4	Geoserver Deployment on Tomcat	10
2.4.1	Obtaining GeoServer and Aggregate Extension	10
2.4.2	Installing Extension	11
2.4.3	Deploying GeoServer	11
3	Deployment	13
3.1	Add Source	13
3.2	Setup Style	15
3.3	Adding Layer	15
3.4	Viewing and Using Layer	17
4	Development	19
4.1	Code Development	19
4.1.1	Creating Pre-Aggregate Index from Source	19
4.1.2	Filtering Words for Nominal Axis	20
4.2	Geoserver Visualization	21
4.2.1	Symbolizers	21
4.2.2	Filters	21
4.2.3	Additional Options	23
5	Running Example	25
5.1	Requirements	25
5.2	Example Table Setup	25
5.2.1	Creating the database	26
5.2.2	Creating the Pre-Aggregate Index	26
5.3	GeoServer Setup	27
5.3.1	Add Source	27
5.3.2	Import Style	29
5.3.3	Create Layer	29
5.3.4	View Layer	30

1 Introduction

Geographical data are typically visualized using various information layers that are displayed over a map. Interactive exploration by zooming and panning actions needs real-time re-calculation. A common operation in calculating with multidimensional data is the computation of aggregates. For layers containing aggregated information derived from voluminous data sets (see for example Figure 1), such real-time exploration is impossible using standard database technology. Calculations require too much time.

The University of Twente has developed “*Stairwalker*”: database technology that accurately aggregates data so that they can geographically be explored in real-time. The technology is a plug-in to common open source technology.

Its core is the *pre-aggregate index*: a database index that cleverly pre-calculates aggregation values such that it can obtain exact aggregation results from voluminous data with high performance. A fast calculation allows to fully recalculate the result for even the slightest movement of the map, such as a panning or zooming action, without loss of accuracy. Thanks to this indexing mechanism, we can provide a scalable real-time calculation: an order of magnitude larger dataset requires only one additional aggregation level.

In geo data visualization, the ability to quickly develop new information layers is important. Although many solutions exist, there is a niche: the combination of visualizing aggregation information, interactive data exploration in real-time, Big Data, calculating exact numbers instead of approximations, and doing so with common open source technology. Our technology for the first time integrates all these features.

Our research partners are the companies Arcadis and Nspyre. They both have struggled with this combination of requirements in many of their projects. Our database index technology is not specific to geographical data. It can be used with all types of multidimensional data. Visualization in business intelligence or eScience can also benefit from it.

Nice to know The company Arcadis developed an application for the DCMR Milieudienst Rijnmond based on the Stairwalk technology to investigate whether people send tweets about unpleasant odors as a possible signal of danger (see Figure 2). This turns out not to be the case, probably because people think that nobody reads the tweets anyway. But if people have the idea that their complaining tweets are read, then tweets might be much more convenient than the reporting of unpleasant odors by telephone.

This manual This manual explains how to use Stairwalker. We first explain in Section 2 how to install the required components in order to have a basic running system. We then explain in Section 3 how to add databases and different kinds of datatypes to Geoserver, an open source server for sharing geospatial data.¹ It is explained how to show and customize layers and views, but also

¹<http://geoserver.org>

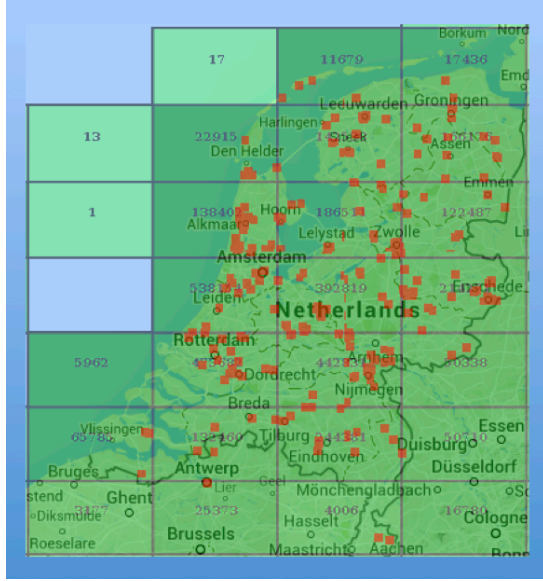


Figure 1: Twitter hotspot detection in The Netherlands, using a coarse grid. The numbers inside the grid cells on the map require an aggregation operation in the database.

how to adjust the system, for example, how to add dimensions or use different dimension types such as median. Finally, Section 4 explains how to extend the system.

Acknowledgements This publication was supported by the Dutch national program COMMIT/.

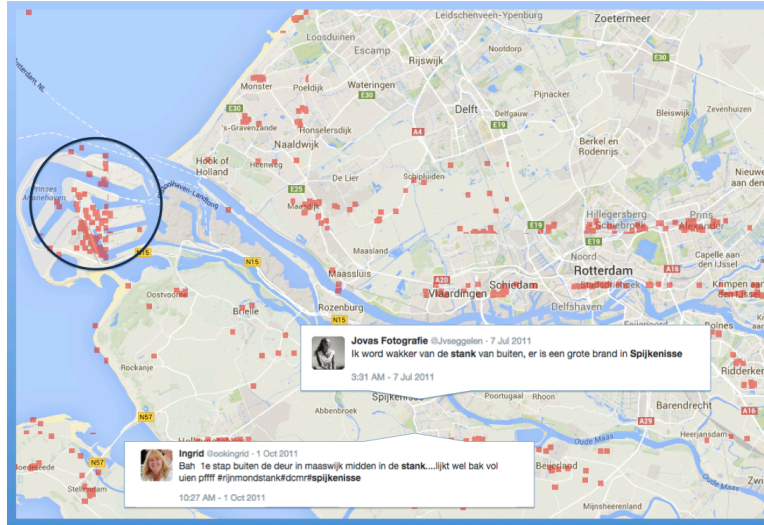


Figure 2: Tweets about excessive smells in the vicinity of the Rotterdam harbor

2 Set Up

In this section a detailed description is given about setting up all the required peripheral programs to use Stairwalker.

2.1 Database Setup

Currently the Stairwalker program only works with the PostgreSQL² and MonetDB³ database management systems. This manual will only concern itself with PostgreSQL. Any further reference to a database implies a PostgreSQL database.

In this section, we give a walk through explaining the steps needed to setup PostgreSQL along with PostGIS⁴ and the database extension for Stairwalker on an Ubuntu Linux operating system.

2.1.1 PostgreSQL Database Manager Setup

Installing Postgres on Linux should be straightforward. Search and install the desired version of PostgreSQL. Below the commands used to search and install version 9.1 of Postgres are shown.

1. `aptitude search postgresql`
2. `apt-get install postgresql-9.1`

²PostgreSQL: <http://www.postgresql.org>

³MonetDB: <http://www.monetdb.org>

⁴PostGIS: <http://www.postgis.net>

For further information or if there are difficulties with the installation more details can be found on the installation page of the PostgreSQL website.

2.1.2 PostgreSQL Configuration

Once Postgres is installed on Linux two alterations will need to be made to the configuration files so that the database can be accessed from outside and by other users.

First, in the `Postgresql.conf` file the `listen_addresses` need to be changed from `localhost` to `all`. Assuming version 9.1 of Postgres this can be done as follows.

1. `cd /etc/postgresql/9.1/main`
2. `vi postgresql.conf`
3. *change* `listen_addresses = 'localhost'` *to* `listen_addresses = '*'`
4. *save and close*

Secondly in the file `pg_hba.conf` a line needs to be added to allow other users in Linux to access Postgres. This can be done as follows:

5. `vi pg_hba.conf`
6. *Add* `host all all 0.0.0.0/0 password`
7. *save and close*
8. `/etc/init.d/postgresql restart`

It should now be possible to create a database user and database in PostgreSQL. In the sequel, we assume a database `'geonames'` is created and used. More information about how this is done can be found on the PostgreSQL manuals webpage.

2.1.3 PostGIS Configuration in PostgreSQL

The next step is to extend PostgreSQL with the PostGIS database expansion. This again should be straightforward: first search for PostGIS and then choose the version that goes with PostgreSQL to install. Note in order to do the install root privileges are required. Below the commands to search and install PostGIS for PostgreSQL-9.1 are shown.

1. `aptitude search postgis`
2. `apt-get install postgresql-9.1-postgis`

More information about installing PostGIS can be found on the PostGIS website.

Once PostGIS is installed some extra configuration is required to add PostGIS functionality to the used database (in our case `'geonames'`). This needs to be done as postgres user. The commands are as follows.

1. `createlang plpgsql geonames`

2. `psql -f 'find/usr/share/postgresql/ -name postgis.sql -print' -d geonames`
3. `psql -f 'find/usr/share/postgresql/ -name spatial_ref_sys.sql -print' -d geonames`
4. `psql -f 'find/usr/share/postgresql/ -name postgis_comments.sql -print' -d geonames`

2.1.4 Serverside Stairwalker Extension in PostgreSQL

Stairwalker is a database extension written in C. The extension has to be compiled and installed in the PostgreSQL database. The extension can be found in the directory `neogeo/pre-aggregate/src/db-extensions/postgres/pa_grid`.

The extension can be installed on Linux using the following commands.

1. *go to the pa_grid directory*
2. `make`
this creates the dynamic library
3. `make install`
this installs the library in the PostgreSQL installation
4. `make sql`
declare the module in the desired database

Note the extension has to be installed specifically for the database which will be used for pre-aggregation. In the makefile (also in the `pa_grid` folder) there is a `DATABASE` macro which should be set to the desired database.

2.2 Pre-Aggregate Database Table

2.2.1 Description of Process

The indexing principle is as follows. For the indexing of the data, several grids are defined, with varying grid sizes. These grids can have any number of dimensions. In the illustration of Figure 3, we show an example for the two-dimensional case. The third dimension in the illustration depicts the aggregation grids at the various levels. For each grid cell, the aggregation function is pre-calculated. Any suitable aggregation function can be used, such as a summation, count, maximum/minimum, etc. When a request comes in, the index returns the pre-calculated values for all entirely included cells. The relatively small amount of non-indexed values to complete the area is the calculated on-the-fly, as illustrated on the right.

Concretely, take the dataset of tweets sent within the Netherlands (see Figure 1). A pre-aggregation of this data, i.e., counts of tweets, could be in the following form. Dimensions are the x- and y-coordinate dimensions. We set a highest granularity (zoom-level). The pre-aggregate algorithm creates blocks bounded by the x- and y-coordinates at the highest granularity and for all those blocks calculates the number of tweets within the bounding box. Each subsequent layer is built up from the previous layer. The final result is a dataset

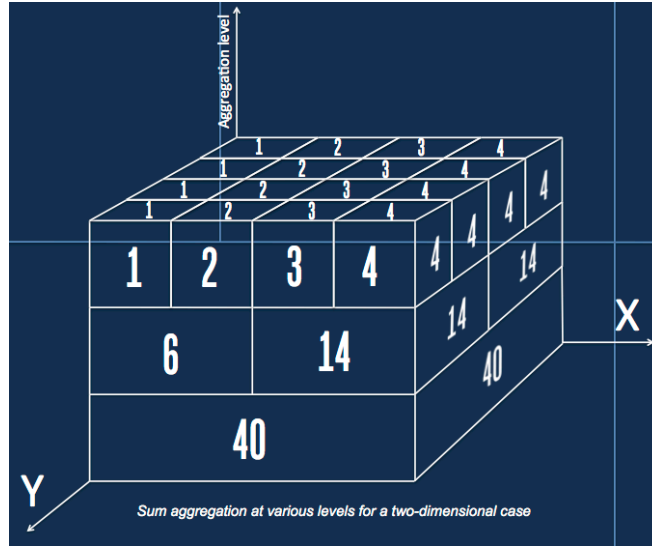


Figure 3: Illustration of the grids concept

which contains the number of tweets in a all bounding boxes at all levels of granularity.

2.2.2 PreAggregate Tool

For off-line creation of the pre-aggregation index, a tool has been developed. The tool can be found in the directory `neogeo/pre-aggregate-tools/`. To generate the binary tools from source, the Appassembler plugin of Maven is used. Run the following command to generate the tool.

```
mvn package appassembler:assemble
```

After successful completion of this command, a new directory `appassembler` has been created in the `target` directory containing a `repo` and a `bin` directory. The `bin` directory contains the actual binaries of the tool (in both Linux/Unix and Windows version) and the `repo` directory contains the tool dependencies. The tool should now be ready for use. An example of how to compile and use the tool to create a pre-aggregation index is given in Section 5.2.2.

The tool is used to create a pre-aggregate index for a table with n dimensions and a measure/aggregate column. It uses with the following commands:

```
usage: create-pa-index
  -axistosplit <axis index>    index of axis to split
  -chunksize <size>            maximum chunk size after split
  -config <file>               PreAggregate XML config file
  -d,--database <dbname>       name of database
  -dbtype <postgresql|monetdb> type of database
```


<code>-h,--host <host></code>	database host name or ip address	7
<code>-help</code>	prints this help message	8
<code>-p,--port <port></code>	port number of the database	9
<code>-password <password></code>	database password	10
<code>-s,--schema <schema></code>	schema name in the database	11
<code>-u,--user <user></code>	database username	12
<code>-v,--verbose</code>	Enable verbose output logging	13

The tool depends on the `PreAggregate.XML` config file which is used to define the PreAggregate index by specifying the column to aggregate, the type of aggregate that should be done and the dimensions to include. In the `neogeo/pre-aggregate-tools/` directory a sample configuration file is included.

See Section 4.1 for more detail about creating a pre-aggregate index not using the above tool.

Apart from creating a new pre-aggregate index table `<original-table-name>_pa`, pre-aggregation also creates/updates two other tables: `pre_aggregate` and `pre_aggregate_axis`. These two tables are support tables for the aggregation. They contain information about which tables have been aggregated and which axis have been used in pre-aggregation.

2.2.3 Current Status Support of Axis Types

The dimensions are referred to in the code as `AggregateAxis`. The axes have a base size and all layers above the base size are built up by aggregating lower layer blocks. Currently the `AggregateAxis` can be split into two different subtypes. The first is a `MetricAxis` and the second is a `NominalAxis`. Each type will be briefly discussed below with examples of data types.

MetricAxis A `MetricAxis` is the default axis type, it supports a continuous data type. Examples of such data types are coordinates and time.

NominalAxis A `NominalAxis` can be used for non-continuous dimensions. An example for such a data type is a word filter. A `NominalAxis` can be used to split the data on occurrence of x predefined words.

For example, if a word filter `{dog,cat,mouse,horse,NOFILTER}` is used and a data block represents some text with the sentence ‘it’s raining cats and dogs’, then three splits would be made. One on the word `dog`, the second on `cat` and the third on `NOFILTER`.

2.2.4 Current Status Support of Aggregation Types

Currently there are 4 different aggregation types which can be used. These are discussed below. Note that there is an `ALL` option which returns all aggregate types in the pre-aggregate index. Furthermore these types can be used to create other types such as average.

The current options are as follows:

1. **ALL**: Returns all the aggregation types mentioned below.
2. **COUNT**: Returns the count of items in an aggregate box.
3. **SUM**: returns the sum of items in an aggregate box. For instance, a sum on the tweet length results in the total amount of characters tweeted in an aggregate box.
4. **MIN**: Returns the lowest value in an aggregate box. With the example of tweet length, it returns the length of the shortest tweet in the aggregate box.
5. **MAX**: Returns the highest value in an aggregate box. With the example of tweet length, it returns the longest tweet in the aggregate box.

It is important to choose the right type in order to get the desired representation of the data. To show the total number of tweets in a tile, one should chose **COUNT** as this will returns the total number of data points in a tile. Whereas if one wants only the highest value in the data (for example the highest building in an area) then **MAX** should be used.

2.3 Tomcat Server Setup

To visually display an aggregated dataset, a third party application is used. This application is called GeoServer, GeoServer is an open source server which can be used to display geospatial data. In order to run GeoServer a web server is needed.

Apache Tomcat⁵ has been used for this purpose. The use of Tomcat is not required for Stairwalker, alternative options may also be used as long as it is possible to deploy WAR files. Whenever web hosting is mentioned in this manual it will be assumed Tomcat is being used.

Information about setting up a Tomcat server can be found on the [Apache Tomcat website](http://tomcat.apache.org/).

2.4 Geoserver Deployment on Tomcat

2.4.1 Obtaining GeoServer and Aggregate Extension

GeoServer⁶ is used to visually display collected data pre-aggregated with the pre-aggregate index. An extension has been written for GeoServer which needs to be included in the web application. This is done by including the JAR files of the extension in the **WEB-INF** of the GeoServer WAR file. A step by step guide is given below in Section 2.4.2. However before following these instructions the necessary JAR and WAR files need to be obtained.

The GeoServer WAR file can be downloaded from the [GeoServer website](http://www.geoserver.org). The custom Java extension files should be built using the source code. The

⁵<http://tomcat.apache.org/>

⁶GeoServer: <http://www.geoserver.org>

extension consists of the `pre-aggregate` and `geoserver-ext` projects in the `neogeo` project. Note that first the JAR file of `pre-aggregate` should be created as it is a dependency of `geoserver-ext`. The JARs can be built using the command line or in an IDE such as Eclipse or Netbeans: using the `clean` and `build` command on a project in Netbeans should build `<project-name>-0.0.1-SNAPSHOT.jar` which can be found in the `target` directory of the respective project.

2.4.2 Installing Extension

The following instructions assume the geoserver extension files are located in the directory `/data/upload/` and the `geoserver.war` file is located in the directory `/data/tmp/tmp_war`. If this is not the case, the file paths in the instructions below should be changed accordingly.

Unpack the WAR file

1. `jar -xvf geoserver.war`

Copy the JAR files of the extension into `WEB-INF/lib` directory of the GeoServer unpacked WAR file

2. `cp /data/upload/pre-aggregate-0.0.1-SNAPSHOT.jar /data/tmp/tmp_war/WEB-INF/lib`
3. `cp /data/upload/geoserver-ext-0.0.1-SNAPSHOT.jar /data/tmp/tmp_war/WEB-INF/lib`

Recreate the GeoServer WAR file

4. `jar -cvf geoserver.war META-INF/ WEB-INF/ index.html data/`

The GeoServer WAR with the stairwalker extension should now be ready for deployment.

2.4.3 Deploying GeoServer

After the GeoServer WAR file has been repacked with the aggregate extension included, it is ready to be deployed in a web server. Section 2.3 describes how to set up a Tomcat web server. Once the server is running, it can be accessed locally with `http://localhost:8080/` assuming default installation configuration were used, otherwise the port number might be different. In the case that the Tomcat server is installed on a different machine, the web server can be accessed by replacing `localhost` with the name or IP address of that machine.

From the Tomcat homepage, it should be possible to access the Tomcat manager webapp. With the default setup it should be possible to login with the following credentials:

username: manager
password: tcmanager

In the Tomcat manager webapp under the **deploy** section, it is possible to upload a WAR file to be deployed. Select the repacked WAR file from Section 2.4.2 and deploy the application. Once the application is deployed it will be displayed in the **application** section of the Tomcat manager webapp. From there, it is possible to follow the path given for the GeoServer application or, if the default configuration was used to go to `http://<Tomcat-IPaddress>:8080/geoserver/web/`.

3 Deployment

This section gives a detailed description of how to import an aggregated database table into GeoServer to get a visual representation of the dataset. First instructions will be given on how to link the database table to GeoServer. Next, creating styles and layers for data representation will be discussed. The final section discusses how to view the data using GeoServer. For this section it is assumed that all the initial preparation discussed in Section 2 has been completed.

GeoServer should already be deployed on a web server (see Section 2.4.3), and can then be accessed with `http://<Tomcat-IPaddress>:8080/geoserver/web/`. It is required to login in to the GeoServer web administration interface. When using the default setup of GeoServer the login credentials are:

username: admin

password: geoserver

A concrete example of how to fully deploy a pre-aggregated index can be found in Section 5.3.

3.1 Add Source

Once logged in to the web administration interface, it is possible to add a new data store to GeoServer. Below are instructions of how to add a new **NeoGeo Aggregate** vector data source which contains the aggregate index created in Section 2.2.2 to the stores in GeoServer.

1. Navigate to **Stores** by clicking on **Stores** link under the **Data** section in the navigator on the left hand side of the web administration interface homepage.
2. On the **Stores** page select the option **Add new Store** located at the top of the page. This leads to a page titled **New Store chooser**.
3. In the list of **Vector Data Sources** the option **NeoGeo Aggregate** should be present, choose this format for the data source.

If the option **NeoGeo Aggregate** is not available, it means the GeoServer extension from Section 2.4.2 was not done correctly.

4. Clicking **NeoGeo Aggregate** will open a new page titled **New Vector Data Source** in which several fields have to be filled out, explanation of mandatory fields can be found in the list below.
5. For an express setup the fields which have already been filled can remain the same.
6. Once all the required fields are filled, click the **Save** button.
7. A new **NeoGeo Aggregate** source is now created and can be viewed and edited in **Sources**.

8. After saving, GeoServer opens the page **New Layer** on which new layers can be created using the **Data Source**. How this is done is discussed in Section 3.3.

Below a list is presented with all mandatory fields on the **New Vector Data Source** page with explanation.

- **Data Source Name** - An arbitrary name which will be assigned to the store.
- **Database type** - The type of underlying database, either PostgreSQL or MonetDB.
- **Hostname** - Hostname of the database server where the aggregation index is maintained.
- **Port** - Port number of the database.
- **Schema** - Name of the schema where the aggregation index is maintained.
- **database** - Name of the database where the aggregation index is maintained.
- **Username** - Username of the used database.
- **Password** - Password of the used database.
- **xSize, ySize, timeSize** - Specifies the dimensions of the grid which is created for every view of the map to calculate the aggregates per cell. The higher the number of cells the more detailed the information.
- **count, sum, minimum, maximum** - Select the boxes of the aggregates which will be used in the visualization. Note that from these basic aggregates more aggregates such as mean can be derived.
- **Enable server-side Stairwalker** - Selecting this causes the data source to rely on the use of the database plugins to use the Pre-Aggregate index. For performance reasons it is highly recommended to use this option. See Section 2.1.4 for more details.
- **Enable query logging** - Selecting this will turn on the logging of all Pre-Aggregate queries into a separate table called `pre_aggregate_logging`.

3.2 Setup Style

In GeoServer, styles are used to render, or make available, geospatial data. Styles are used to visually represent the aggregation index which is represented in a layer. In GeoServer layers are written in Styled Layer Descriptor (SLD) which is a subset of XML. GeoServer comes setup with several different styles, however, to get the most out of the dataset it is best to develop a style specific to the layer which represents that data.

In this section only instructions on how to add new styles to GeoServer are given. For information on how to edit styles, see Section 4.2 or the GeoServer user manual⁷ which gives an in depth guide on developing styles.

1. Navigate to **Styles** by clicking on the **Styles** link under the **Data** section in the navigator on the left hand side of the web administration interface homepage.
2. On the **Styles** page select the option **Add a new style** located at the top of the page.
3. A new page titled **New Style** should open. There are now two possibilities, either a new style can be developed completely in the browser or a SLD file can be imported.
4. To import an already created SLD file scroll to the bottom of the page and press the **Choose File** button.
5. Select the style which should be imported and then press **Upload...** in the browser.
6. This fills in the **Name** field with the name of the file and the SLD editor with the content of the file.
7. It is possible to check the syntax of the SLD code by pressing the **Validate** button at the bottom of the page. At the top the page GeoServer will give feedback on the SLD code, either error messages or a no validation errors message.
8. Finally the style can be saved by pressing the **Submit** button at the bottom of the page.

3.3 Adding Layer

In GeoServer, a layer refers to raster or vector data that contains geographic features. Layers represent each feature (axis in the pre-aggregate index) of a dataset that needs to be represented. All layers much have a source of the data which in this case was setup in Section 3.1. More information about layers can be found in the GeoServer User Manual.

Creating a layer for a pre-aggregate index dataset can be done as follows:

⁷<http://docs.geoserver.org/stable/en/user/styling/index.html#styling>

1. Navigate to **Layers** by clicking on the **Layers** link under the **Data** section in the navigator on the left hand side of the web administration interface homepage.
2. On the **Layers** page select the option **Add a new resource** located at the top of the page.
3. This leads to a new page where the **Store** which contains the layer needs be chosen from a drop-down list. If there are no **Stores** available make sure one was added, see Section 3.1
4. Choose the **Store** in which the aggregation index is stored.
5. Once a **Store** is selected a list of resources contained in the **Store** is given. These resources are the different aggregated indexes in the database which was linked to a **Store** in Section 3.1.
6. Select the pre-aggregate index which should be visualized in a layer by clicking the **Publish** link corresponding to the **Layer name** of the aggregate index.

At this point a layer has been selected to be published. This layer will be a visual representation of the data from the aggregation index created in Section 2.2.2. In order to make sure the correct geographical location is used in GeoServer and to give the layer a fitting style the following steps have to be taken in on the **Edit Layer** page.

7. In the **Data** tab the following sections and fields should be filled in.
 - (a) In the **Basic Resource Info** there are some labeling fields. Standard the **Name** and **Title** are `<aggregation-index-tablename>` followed by `___myAggregate`. These can both changed to whatever is desired. However make sure the that the **Enabled** box is ticked in this section.
 - (b) In the **Coordinate Reference Systems** section there are three fields.
 - i. **Native SRS** should be `EPSG:4326`.
 - ii. **Declared SRS** should be `ESPG:3857`. This coordinate system is used since it is what is usually used for tile based map representation.
 - iii. **SRS handling** should be `Reproject native to declared`.
 - (c) In the **Bounding Boxes** the coordinates corresponding to the data from the aggregation index is calculated for GeoServer. For **Native Bounding Box** click **Compute from data** and for **Lat/Lon Bounding Box** click **Compute from native bounds**.
 - (d) All other sections in this tab are of little importance in a basic deployment.

8. Next in the **Publishing** tab a style can be added to the layer, the default style of a layer is **polygon**. In the section **WMS Settings** the field **Default Style** can be changed by selecting the desired style from the drop-down menu. For more about styles and creating styles see Section 3.2 and Section 4.2.

If the aggregation index does not contain a time dimension the setup of the layer is now complete and can be saved. However if the aggregation index does have a time dimension some additional adjustments need to be made which are described below.

9. Select the **Dimensions** tab.
 - (a) Enable the the **Time** dimension.
 - (b) As **Attribute** select **starttime**.
 - (c) Do not set an **End Attribute**.
 - (d) As **Presentation** select **Continuous interval**.
10. Save the layer.

The layer which represents the dataset with a style created in Section 3.2 has now been created and is ready for use. A layer can be edited once it has been created so if changes need to be made a new layer should not be created. Section 3.4 shows how to preview a layer and Section ?? discusses how to use a GeoServer layer with OpenLayers⁸ to a geospatial visualization of the data on a web page.

3.4 Viewing and Using Layer

In GeoServer it is possible to get a preview of layer such as the one created in Section 3.3. Previewing a layer can be done as follows:

1. Navigate to **Layer Preview** by clicking **Layer Preview** link under the **Data** section in the navigator on the left hand side of the web administration interface homepage.
2. The **Layer Preview** page will have a list of all configured layers with can be previewed in various formats.
3. Locate the layer which should be shown and from the **All Formats** column choice any **WMS** format.
4. After selecting a format to view the layer a new page will open with a visual representation of the top most layer of dataset.

⁸<http://openlayers.org/>

Note that other preview formats should also be possible. For example it is possible to use the OpenLayers preview format which allows one to navigate the geospatial data. In section Section ?? OpenLayers will also be used to visualize the dataset on a web page. One drawback is that for every movement in the preview a new query has to be calculated. When server side stairwalker (Section 2.1.4) is not setup calculating can be time consuming. Therefore during development of the pre-aggregation index and testing it is advised to use a static preview and only once everything works as desired to use a dynamic preview.

If the layer contains a nominal axis it is possible to alter the value of the nominal with which the data is filtered. This is done by adding a parameter in the request made to the GeoServer extension. By extending the HTML request sent to GeoServer with `&VIEWPARAMS=<TYPE>:<VALUE>;` the nominal axis filter is used. Currently the extension only supports the nominal type **keyword**. If the pre-aggregate index was created using a nominal axis (splitting on **VALUES**) then using `&VIEWPARAMS` will split the visualization on a give **VALUE**. `&VIEWPARAMS` can accept more `<TYPE>:<VALUE>;` tuples, however parsing type will need to be extended on in the code. For more information see Section 4.1.2.

It may be that a preview fails to load, this can be due to two reasons. The first is an error in the style, in this case the style needs to be tested which can be done by validating the style in GeoServer. For more information about styles see Section 4.2. The second reason is an error is creating an SQL query for the pre-aggregated index. If no mistakes were made during setup and in creating a pre-aggregate index there will be a need to dive into the code where detailed logging is done.

4 Development

4.1 Code Development

In this section, some pointers will be given to important sections of code in terms of processing pre-aggregate indexes which are used by the GeoServer extension.

Using the source code GeoServer and running it from an IDE, provides useful information for troubleshooting, as logging information is printed directly to the console, and for constant redeployment a new WAR file does not have to be remade every time. The source code of the Stairwalker project is available on GitHub⁹. The two important packages from this project are **geoserver-ext** and **pre-aggregate**. The first is the extension used in GeoServer to handle pre-aggregation indexes as data sources and the second is the package which is used to create the pre-aggregate index.

4.1.1 Creating Pre-Aggregate Index from Source

Creating a pre-aggregate index can be done using a tool as shown in Section 2.2.2. It is also possible to create a new method in the **Test.java** of **pre-aggregate** in which the steps for creating the pre-aggregation index can be done manually. In the **main** of this class configuration file **database.properties** is read, which contains login information about the database which contains the dataset, and connection is made to said database. Next a pre-aggregate index is made using the custom made method which describes the pre-aggregation. How to set up such a method is described in this subsection.

Firstly for each axis on which the dataset will be split an **AggregateAxis** variable will be defined. Next these variables should be initiated with a axis type (**metric** or **nominal**), in the case of **nominal** also a word list (which contains the words on which the dataset will be split) should be created. The constructors of both axis types are as follows:

```
public MetricAxis(String columnExpression, 1
                  String type, Object BASEBLOCKSIZE, short N); 2
3
public NominalAxis(String word_collection_column, 4
                   String word_index_column, String wordlist_str, 5
                   String name); 6
```

Once all axes have been initialized, an array containing all axes should be created which will be used as input for the **PreAggregate** variable created in the next. The next step is to create a pre-aggregate index by creating and initiating a **PreAggregate** variable. The constructor of **PreAggregate** is as follows:

```
public PreAggregate(Connection c, String schema, 1
                    String table, String override_name, 2
                    String label, AggregateAxis axis[], 3
                    String aggregateColumn, String aggregateType, 4
```

⁹<https://github.com/utwente-db/neogeo>

```
int aggregateMask, int axisToSplit, 5
long chunkSize, Object[][] newRange); 6
```

Afterwards the connection to the database should be closed. A method containing these components is created and it can be statically run in the main. Note that with the `NominalAxis`, some preprocessing maybe required on the dataset. In order to do this, the help method `tagWordIds2Table` is used. On `NominalAxis` the `tagWordIds2Table` method has the following arguments:

```
public void tagWordIds2Table(Connection c, 1
                             String schema, String org_table, 2
                             String new_table); 3
```

4.1.2 Filtering Words for Nominal Axis

In the package `geoserver-ext`, specifically the class `AggregationFeatureSource` creates the SQL query which requests data from the pre-aggregated index for which GeoServer is built. If the pre-aggregated index contains a `Nominal` axis, then it is possible to pass along words to filter the axis in GeoServer using `VIEWPARAMS` (see Section 3.4). In order to include this filter parameter, the method `getReaderInternal` and possibly `reformulateQuery` should be extended.

In the method `getReaderInternal`, the layer request sent to GeoServer is parsed including the `VIEWPARAMS`, so a variable should be created in the method which matches the word which should be filtered given by `<TYPE>:<VALUE>`. This variable can then be passed to the `reformulateQuery` method. It is then possible in `reformulateQuery` for a specific `AggregateAxis` to split on the variable passed along in `getReaderInternal`.

4.2 Geoserver Visualization

This section discusses most of the possibilities that GeoServer has to offer when it comes to the visualization of the data. The GeoServer manual has some information on this subject, which can be found on: <http://docs.geoserver.org/2.5.x/en/user/styling/sld-reference/index.html>.

In the following, more information on what the differences are between the certain options and some ways to implement these options. This section uses some of the examples used in our demo to give some more insight what can be done and hopefully making it easier to realize what is desired. Most information can be found on the website, below a discussion is given on what types are best used when, but also some more information on how to make them work properly.

4.2.1 Symbolizers

In SLD, there are three different symbolizers, a linesymbolizer, a pointsymbolizer and a polygonsymbolizer. A pointsymbolizer is used when the data that has to be represented is best shown as points. It does exactly what it says, you'll get a map with points on it and each point will represent a data-object from your dataset. This symbolizer can be really handy in certain situations. For example when you want to show all locations where a rare species of a bird has been found, it will show a map with all points where a bird has been reported.

A linesymbolizer is used when the data that has to be displayed is best shown in lines. This symbolizer is best used to display roads for example. It isn't a symbolizer that can be used to represent data very well, but it is more used for pre-defined data, such as rivers, roads etc.

A polygonsymbolizer is used when the data that you want to represent has to be displayed in two-dimensional objects. There are many possibilities in a polygonsymbolizer. It is possible to make a simple square but it also has the option to make circles or triangles. It is one of the most commonly used symbolizers. A good example where a polygonsymbolizer is used, is to display the amount of people living in cities. This can be done with a circle polygon and that the circle will get bigger when more people live in a city.

4.2.2 Filters

Filters are the most important function when it comes to making a custom style. A filter is basically the basis of a fancy layer. What a filter does is that it makes a ruling and if that ruling is met, the color, labeling etc will be done. In SLD, it is possible to have an unlimited amount of filters so the possibilities are many. The following filter expression can be used:

- `<PropertyIsEqualTo>`
- `<PropertyIsNotEqualTo>`
- `<PropertyIsLessThan>`

- `<PropertyIsLessThanOrEqualTo>`
- `<PropertyIsGreaterThan>`
- `<PropertyIsGreaterThanOrEqualTo>`

An example on how a single filter can be used is the following:

```
<ogc:Filter>
  <ogc:PropertyIsLessThan>
    <ogc:PropertyName>testvalue</ogc:PropertyName>
    <ogc:Literal>200</ogc:Literal>
  </ogc:PropertyIsLessThan>
</ogc:Filter>
```

This example tests if `testvalue` is less than 200. If this is the case, one can specify what the filter should do. Below is the complete example that does something with this filters.

```
<Rule>
  <Name>SmallPop</Name>
  <Title>Less Than 100</Title>
  <ogc:Filter>
    <ogc:PropertyIsLessThan>
      <ogc:PropertyName>testvalue</ogc:PropertyName>
      <ogc:Literal>100</ogc:Literal>
    </ogc:PropertyIsLessThan>
  </ogc:Filter>
  <PolygonSymbolizer>
    <Fill>
      <CssParameter name="fill">#38FF19
    </CssParameter>
      <CssParameter name="fill-opacity">1.0
    </CssParameter>
    </Fill>
    <Stroke>
      <CssParameter name="stroke">#000000
    </CssParameter>
      <CssParameter name="stroke-width">1.0
    </CssParameter>
    </Stroke>
  </PolygonSymbolizer>
</Rule>
```

What this example does is that if the `testvalue` is below 100, it will fill a polygon with the color: `#38FF19`. If this is not the case, it will go to the next rule (if there is one; otherwise it does nothing). The image shows a graph of the implementation we made for our data. The image has different kind of colors for the amount of data in a tile. If the amount is high, the color will become more red, and if there is little data, the tile will be green.

4.2.3 Additional Options

GeoServer SLD has a lot of options when it comes to customizing the data display that you've made. Below are some of the important features that are commonly used in GeoServer.

Halo: A halo gives a glow behind the current label. It should always be used in a textsymbolizer, since this is the only place where you can add a halo. To use a halo is very simple: you include `<Halo></Halo>` and in between it is possible to add `<Radius>` and `<Fill>`. For more information on how to use a `<Fill>` we refer to the GeoServer SLD cookbook.

Anchorpoint: An anchor point is a handy tool to place your label in a possible position. It is used as shown below. Important to note is that you can set where the anchor point is (for example, above the point) and that you can move it based on the anchor point, for instance make it go all the way to left (negative X placement) or all the way to the right (positive X placement)

```
<PointPlacement>
  <AnchorPoint>
    <AnchorPointX>0.5</AnchorPointX>
    <AnchorPointY>0.0</AnchorPointY>
  </AnchorPoint>
  <Displacement>
    <DisplacementX>0</DisplacementX>
    <DisplacementY>25</DisplacementY>
  </Displacement>
  <Rotation>-45</Rotation>
</PointPlacement>
```

1
2
3
4
5
6
7
8
9
10
11

Opacity: Opacity is the transparency of either a label, point, polygon or line. It can be used to paint layers over each other (setting Opacity to 0). It is often used in cases where multiple data sets have to be displayed in the same tile (used in our example as well). The way you use it is the following:

```
<Opacity>0.3</Opacity>
```

1

Rotation: Rotation is the function that is used to turn shapes and labels in SLD. It is handy if you want to turn tiles or make labels line up with lines better. The way to use it is simply in the section that has to be rotated just add the following code: `<Rotation>-45</Rotation>` for a negative 45 degree turn.

Graphic Fill: A graphic fill is used in case a picture/image has to be shown in a layer. It has a lot of possibilities since every picture/image can be added in this way. The implementation is a little more complex; below is an example of a graphic fill.

```
<FeatureTypeStyle>
  <Rule>
    <PolygonSymbolizer>
      <Fill>
```

1
2
3
4

```

    <GraphicFill>
      <Graphic>
        <ExternalGraphic>
          <OnlineResource
            xlink:type="simple"
            xlink:href="colorblocks.png" />
          <Format>image/png</Format>
        </ExternalGraphic>
        <Size>93</Size>
      </Graphic>
    </GraphicFill>
  </Fill>
</PolygonSymbolizer>
</Rule>
</FeatureTypeStyle>

```

More options and information can be found in the SLD cookbook on the GeoServer website. This section was meant to give some more insight about commonly used functions.

5 Running Example

In this section, a concrete example is given to show how to use Stairwalker. The example runs from creating a pre-aggregation index from a given dataset to geospatially representing the data using GeoServer. For the example, a dataset is provided in terms of a .csv file. It concerns an exported list of tweets sent in the UK and which carry a GPS coordinate.

The result expected in this example can be seen in Figure 11. Different colors represent the number of tweets in a region. The result only shows the highest granularity. Each tile shows how many tweets are sent from a location within the tile. If this layer is combined with a map, it would be clear that the region of this figure is directly above the UK.

5.1 Requirements

Before showing how to create the given example, all necessary installations should have been completed and the required files obtained. The following programs should be installed:

1. PostgreSQL (Section 2.1.1)
2. PostGIS for a PostgreSQL database (Section 2.1.3)
3. Tomcat or another web service server (Section 2.3)
4. GeoServer with extension (Section 2.4.3)

Also the following files should be at the ready:¹⁰

1. The tool to make a pre-aggregate index: Pre-Aggregate-Index tool (Section 2.2.2)
2. Configuration file for the tool: `runningexample.config.xml`
3. Sample dataset: `RunningExample.csv`
4. Sample SLD file: `RunningExampleSLD.xml`

Furthermore it is useful to have a frontend tool in which the PostgreSQL database can be managed. During development the tool pgAdmin¹¹ was used.

5.2 Example Table Setup

The first step of the process is to have a dataset which to be aggregated. For this example, a dataset is supplied in the form of a .csv file. In order to import this, first create the table in the database. This can be done with the following SQL query:

¹⁰The files can be found in the `RunningExample` directory in the same Git as this manual.

¹¹<http://www.pgadmin.org/>

5.2.1 Creating the database

```
CREATE TABLE runningexample (  
    id_str character varying(25),  
    tweet text,  
    user_name text,  
    place_name text,  
    time timestamp with time zone,  
    reply_to text,  
    place_id bigint,  
    len bigint,  
    coordinates geometry,  
    CONSTRAINT enforce_dims_coordinates CHECK ((  
        st_ndims(coordinates) = 2)),  
    CONSTRAINT enforce_geotype_coordinates CHECK  
        (((geometrytype(coordinates) = 'POINT'::  
            text) OR (coordinates IS NULL))),  
    CONSTRAINT enforce_srid_coordinates CHECK ((  
        st_srid(coordinates) = 4326))  
);
```

Once the table is created, import the .csv file into the table.

Note that `RunningExample.csv` contains column headers, uses ; as column separators and " as quote separators. After the import is done, a pre-aggregate index can be created.

5.2.2 Creating the Pre-Aggregate Index

An in depth discussion of the use of the pre-aggregate index tool is given in Section 2.2.2. In the example, only commands will be given with only brief explanations.

First the pre-aggregate index creation tool needs to be compiled, which can be done with the command below executed in the `pre-aggregate-tools` directory.

```
mvn package appassembler:assemble
```

The next step is to put the pre-aggregate tool config file in the `pre-aggregate-tools` directory. Once this is done, the tool can be called with the following command. Note some variables need to be set in the listing below. These are `<database>`, `<host>`, `<port>`, `<pass>`, `<user>`, which should be filled according to how PostgreSQL was set up.

```
target\appassembler\bin\create-pa-index -config  
    runningexample.config.xml -d <database> -dbtype  
    postgresql -h <host> -p <port> -password <pass> -s  
    public -u <user>
```

This creates a pre-aggregate index of the dataset. In the database three new tables are created: The pre-aggregate index named `runningexample_pa` and



Figure 5: Adding new **Store** to GeoServer

two help tables which keep track of the indexes and the axes used by those indexes. All the work on the side of the database is now done, and the next step is to visualize the dataset using GeoServer.

5.3 GeoServer Setup

This section gives a step by step guide of how to create a visual geospatial representation using the pre-aggregated index of the example dataset. This will be done using GeoServer, specifically the GeoServer web administration interface. This section offers a concrete version of the deployment discussed in Section 3.

Figure 4 shows the **Data** section of the navigator which can be found on the left hand side of the web administration interface. The links in this section will be used to navigate between different pages needed to configure the whole setup.

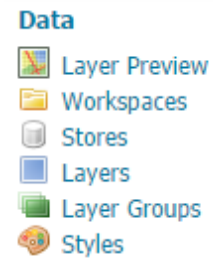


Figure 4: Data section of navigator

5.3.1 Add Source

A data source is added in the following way:

1. Click on the **Stores** link in the **Data** section shown in Figure 4.
2. The **Stores** will open, the top of the page looks like Figure 5(a); click on the **Add new Store** link.
3. A selection of different **Vector Data Sources** is now available. Select **NeoGeo Aggregate** as shown in Figure 5(b).
4. After selecting **NeoGeo Aggregate** as **Vector Data Source** a page like Figure 6(a) will open. Fill in all the fields as shown. Some values may differ depending on how the database is setup. More exact information can be found in Section 3.1.

New Vector Data Source

Add a new vector data source

NeoGeo Aggregate
NeoGeo aggregation index query

Basic Store Info

Workspace *

nurc

Data Source Name *

RunningExampleDataStore

Description

☒ Enabled

Connection Parameters

Database type *

POSTGRES

hostname *

localhost

port *

5432

schema *

public

database *

TestDatabase

Username *

postgres

Password *

Namespace *

http://localhost/nurc

xSize *

10

ySize *

10

timeSize *

1

☒ count

☐ sum

☐ minimum

☐ maximum

☐ enable query logging

Save

Cancel

(a) New Vector Data Source

Edit Layer

Edit layer data and publishing

nurc:RunningExampleLayer

Configure the resource and publishing information for the current layer

Data Publishing Dimensions Tile Caching

Basic Resource Info

Name

RunningExampleLayer

☒ Enabled

☒ Advertised

Title

RunningExampleLayer

Abstract

Keywords

Current Keywords

features

aggregate_runningexample___myAggregate

Remove selected

New Keyword

Vocabulary

Add Keyword

Metadata links

No metadata links so far

Add link | Note only FGDC and TC211 metadata links show up in WMS 1.1.1 capabilities

Coordinate Reference Systems

Native SRS

EPSG:4326

EPSG-WGS 84...

Declared SRS

EPSG:3857

Find...

EPSG-WGS 84 / Pseudo-Mercator...

SRS handling

Reproject native to declared

Compute from data

Compute from native bounds

Compute from data

Compute from native bounds

Compute from data

Compute from native bounds

Compute from data

Compute from native bounds

Compute from data

Compute from native bounds

Compute from data

Compute from native bounds

Compute from data

Compute from native bounds

Compute from data

Compute from native bounds

Compute from data

Compute from native bounds

Compute from data

Compute from native bounds

Compute from data

Compute from native bounds

Compute from data

Compute from native bounds

Compute from data

Compute from native bounds

Compute from data

Compute from native bounds

Compute from data

Compute from native bounds

Compute from data

Compute from native bounds

Compute from data

Compute from native bounds

Compute from data

Compute from native bounds

Compute from data

Compute from native bounds

Compute from data

Compute from native bounds

Compute from data

Compute from native bounds

Save

Cancel

(b) Edit Layer page

Figure 6: Configuring a store and layers

5. Once everything is filled out click the **Save** button. This leads to page where **Layers** can be published. However before that is done, first the **Style** should be imported.

5.3.2 Import Style

Importing a style is done as follows:

6. Click on the **Styles** link in the **Data** section shown in Figure 4.
7. Click on the **Add a new style** button which will go to a page similar to Figure 7 although empty.
8. Import the **RunningExampleSLD.xml** file by using the **Choose File** button. The **Upload...** link is highlighted in red in Figure 7.
9. Once the style has been uploaded, the **New style** page should look like Figure 7.
10. Press the **Save** button.

The style used in this example has been imported in GeoServer and now the layer is ready to published.

5.3.3 Create Layer

Creating a new layer is done as follows:

11. Click on the **Layers** link in the **Data** section shown in Figure 4.
12. This opens the **Layers** page, here click on the **Add a new resource** button. This open a page similar to Figure 9.
13. Select the **Publish** action for the example layer.
14. A page like Figure 6(b) will open. Set highlighted fields to match Figure 6(b). More exact information about these fields can be found in Section 3.3.
15. After the fields in the **Data** are filled in, go to the **Publishing** tab, see Figure 8.
16. Set the default style to **RunningExampleSLD** like in Figure 8.
17. Press the **Save** button.

A layer for the example dataset has now been created and is ready to be viewed.


New style

Type a new SLD definition, or use an existing one as a template, or upload a ready made style from your file system. The editor can provide a valid SLD document.

Name

Workspace

Copy from existing style
 [Copy ...](#)

 **12pt**

```
1 <?xml version="1.0" encoding="ISO-8859-1"?>
2 <StyledLayerDescriptor version="1.0.0"
3 xsi:schemaLocation="http://www.opengis.net/sld StyledLayerDescriptor.xsd"
4 xmlns="http://www.opengis.net/sld"
5 xmlns:ogc="http://www.opengis.net/ogc"
6 xmlns:xlink="http://www.w3.org/1999/xlink"
7 xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
8   <!-- a Named Layer is the basic building block of an SLD document -->
9   <NamedLayer>
10     <Name>polygon_nt_color1</Name>
11     <UserStyle>
12       <!-- Styles can have names, titles and abstracts -->
13       <Title>Default Polygon with color based on cnt 1</Title>
14       <Abstract>A sample style that draws a polygon and adds the feature as text</Abstract>
15       <!-- FeatureTypeStyles describe how to render different features -->
16       <!-- A FeatureTypeStyle for rendering polygons -->
17       <FeatureTypeStyle>
18         <Rule>
19           <Name>LabelPOP</Name>
20           <Title>Less Than 1000000</Title>
21           <ogc:Filter>
22             <ogc:PropertyIsLessThan>
23               <ogc:PropertyName>counttaggr</ogc:PropertyName>
24               <ogc:Literal>1000000</ogc:Literal>
25             </ogc:PropertyIsLessThan>
26           </ogc:Filter>
27           <TextSymbolizer>
28             <Label>
29               <ogc:PropertyName>counttaggr</ogc:PropertyName>
30             </Label>
31             <Font>
32               <CssParameter name="Font-Family">Times</CssParameter>
```

SLD file
 No file chosen

Figure 7: Importing SLD style from file

5.3.4 View Layer

The final GeoServer step is to preview the layer. The preview only shows the highest granularity of the aggregation index. Getting a preview of a layer is done as follows:

18. Click on the **Layer Preview** link in the **Data** section shown in Figure 4.
19. The **Layer Preview** page opens which displays all viewable layers like in Figure 10.
20. A preview format need to be selected from the drop-down menu highlighted in Figure 10.
21. Select a **WMS** preview format such as **PNG**.
22. A new web page will load (this might a few seconds depending on whether or not the server side extension is enabled).

Edit Layer

Edit layer data and publishing

nurc:RunningExampleLayer

Configure the resource and publishing information for the current layer

Data
Publishing
Dimensions
Tile Caching

HTTP Settings

☐ Response Cache Headers

Cache Time (seconds)

WMS Settings

☒ Queryable

☐ Opaque

Default Style

RunningExampleSLD ▾

Less Than 100000000

☐ Less Than 1

☐ Less Than 6

☐ Less Than 100

☐ Less Than 500

☐ 100 to 250

☐ 250 to 2500

☐ 2500+

Figure 8: Adding a Style to the Layer

23. The final result will look like Figure 11.

The layer showing the example dataset is now complete. The values of each square in the layer is calculated using the pre-aggregate index of the dataset. See ?? to learn how the layer can be used in combination with other tools such as OpenLayers¹² to create a dynamic map which updates data on the fly.

¹²<http://openlayers.org/>

New Layer

Add a new layer

You can create a new feature type by manually configuring the attribute names and types. [Create new feature type...](#)
Here is a list of resources contained in the store 'RunningExampleDataStore'. Click on the layer you wish to configure

<< < > >> Results 1 to 1 (out of 1 items)		Search	
Published	Layer name		Action
	aggregate_runningexample__myAggregate		Publish
<< < > >> Results 1 to 1 (out of 1 items)			

Figure 9: Publishing a Layer

Layer Preview

List of all layers configured in GeoServer and provides previews in various formats for each.


<< < > >> Results 1 to 1 (out of 1 items)				Search
Type	Name	Title	Common Formats	All Formats
	nurc:RunningExampleLayer	RunningExampleLayer	OpenLayers KML GML	Select one
<< < > >> Results 1 to 1 (out of 1 items)				

Figure 10: Previewing a Layer

0	0	0	0	0	1	7	1	0	1
0	0	0	0	140	1263	490	0	0	0
0	0	0	2	269	3119	1141	83	0	0
0	0	1	0	5	555	862	33	0	0
1	0	5	0	0	276	16	0	0	0
0	0	0	0	0	0	1	2	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	20	0	0	0
2	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0

Figure 11: Preview of whole dataset