RDF4GeoKettle

RDF IN AN ETL ENVIRONMENT

Current process to generate and publish RDF with geometries support

1. Pre-processing
   - Validation
   - Change of coordinate system
2. Transform the input (SHP, KML, GML) into RDF.
3. Publish all the files into a selected RDF Endpoint.

Current solutions

- Improve existing systems (*TripleGeo* improved geometry2rdf).
- Create a new system for specific tasks (*GeoTriples*, *GeomRDF*).
- Create a new framework (*DataLift* was built as a new framework)

But...

Those are not solutions. Why?

- All current work is wasted.
- Consumes a lot of time in development.
- Requires understanding of all the parts of the new or old system.
- Requires continuous support and development of the new system.
- Creates a new community of **0** users

What is the solution?

New methodology:

- Use existing tools
- Develop components instead of systems

No need to develop a entirely new system.

- Chosen ETL for RDF Generation: GeoKettle
- Chosen Visualizer: Map4RDF

Why GeoKettle?

- It is an ETL environment with geometries support (WKT Format).
- Supports most of geospatial format (SHP, GML, KML, and others).
- Support for data transformation (filters, SRS transformation, fields calculation, and more).
- Support for data publication (KML, GML, Excel, and more).
- 6+ years of development.
- Has an existing community of users.
- Drag & Drop usage.
- Allows use and development of plugins.
- Kettle/Pentaho plugins can be adapted to work with GeoKettle

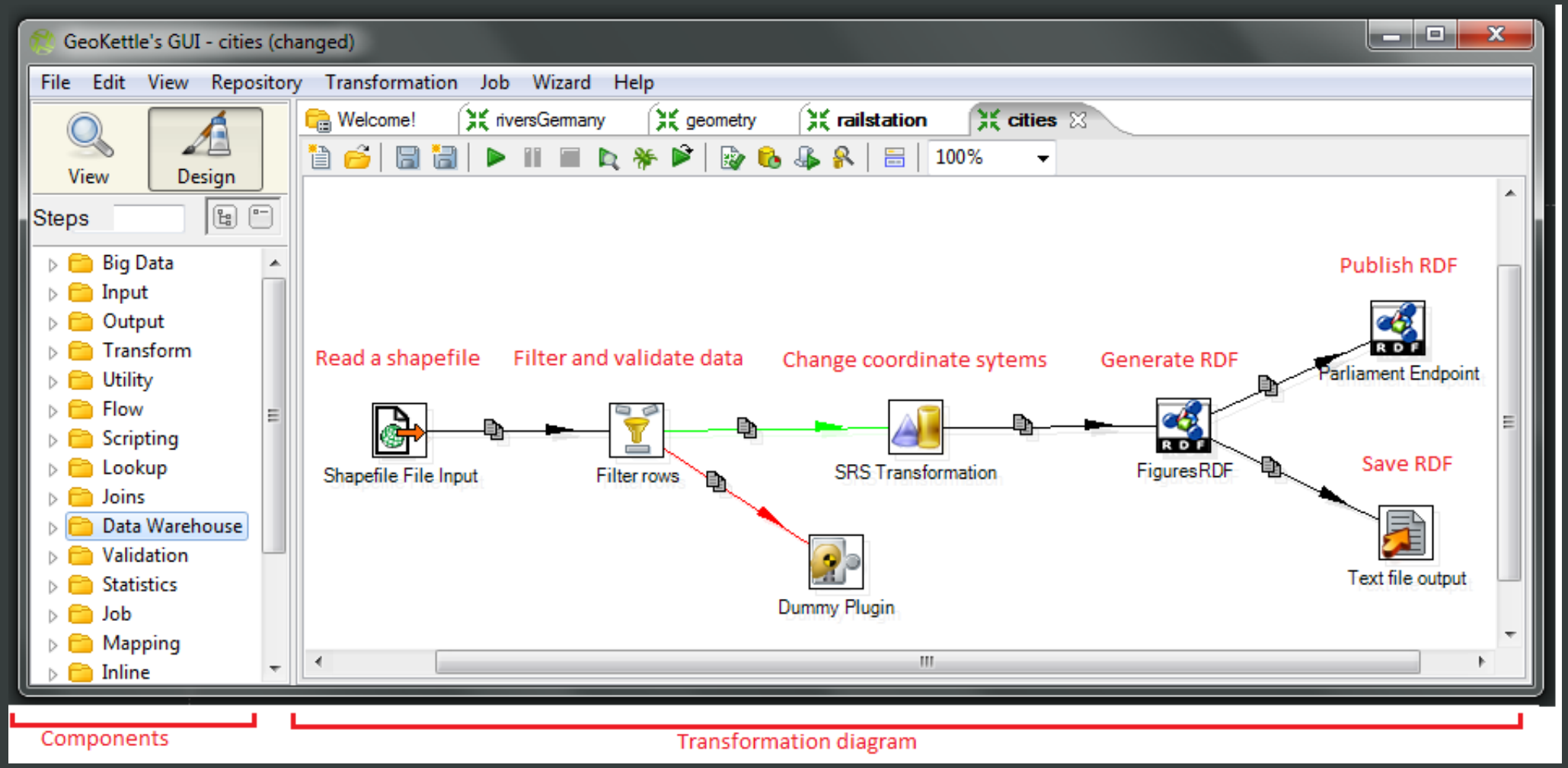
Solution for RDF Generation and publication

GeoKettle with RDF Plugins

New developed plugins:

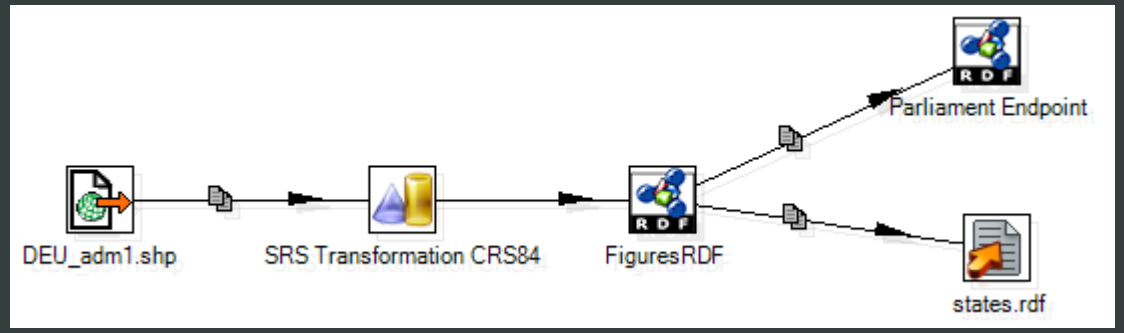
- **RDF Input**: Adaptation of Kettle/Pentaho plugin. Read data from RDF Endpoints using SPARQL.
- **FiguresRDF**: Transform input into RDF (using GeoSPARQL standard).
- **Parliament Endpoint**: Publish RDF in Parliament.
- **Silk Execution Plugin**: Executes Silk using generated RDF.

General view

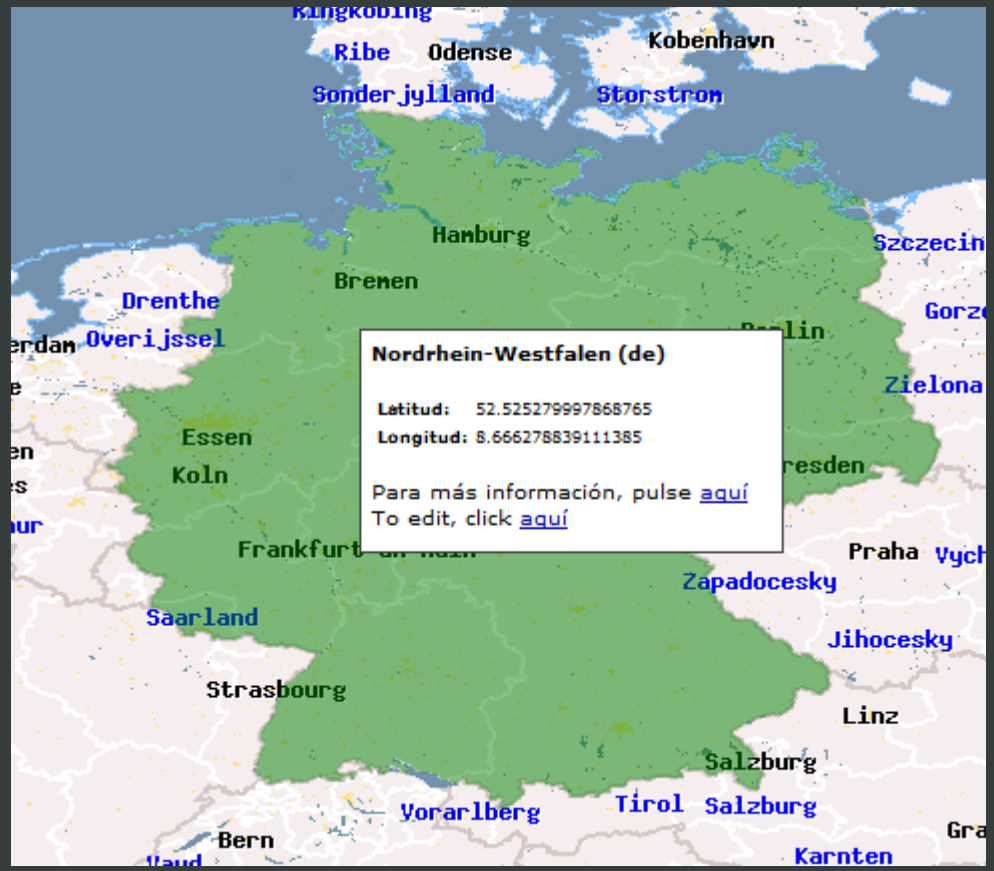


Example of use 1

- States of Germany
- Shapefile from: http://www.gadm.org/download
- Geometry type: Polygons



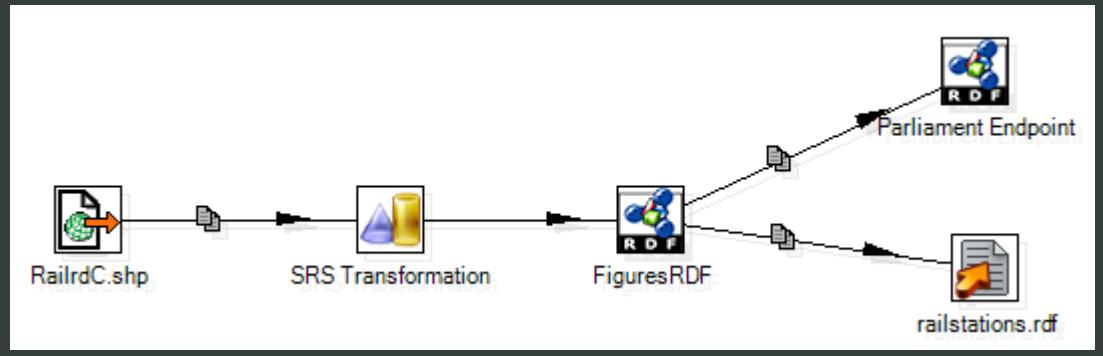
GeoKettle - Generation



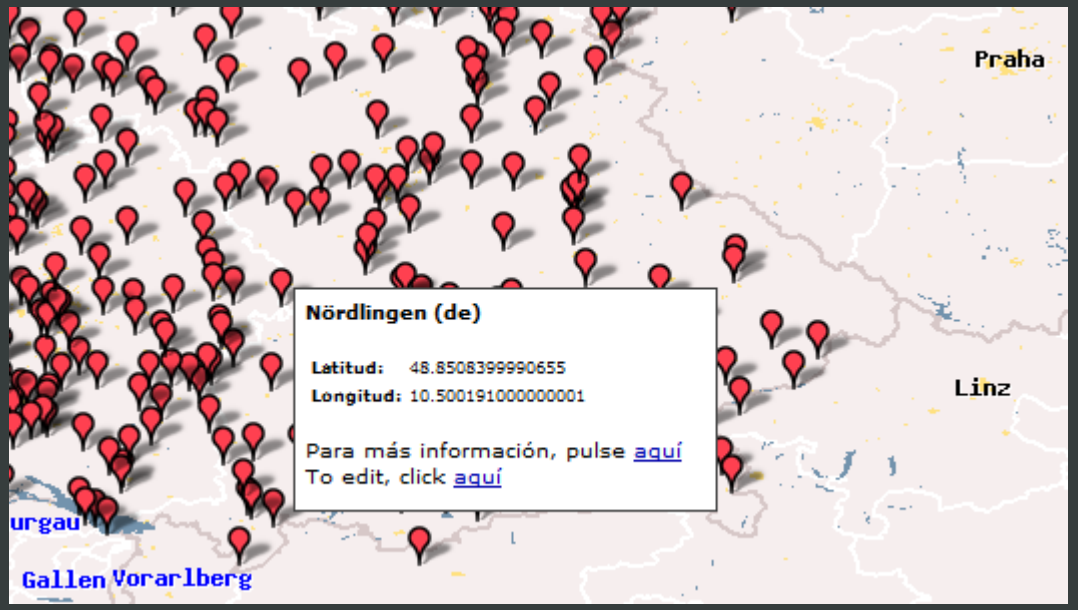
Map4RDF - Visualization

Example of use 2

- Railway stations in Germany
- Shapefile from: http://www.eurogeographics.org/content/euroglobalmap-opendata?sid=7598
- Geometry type: Points



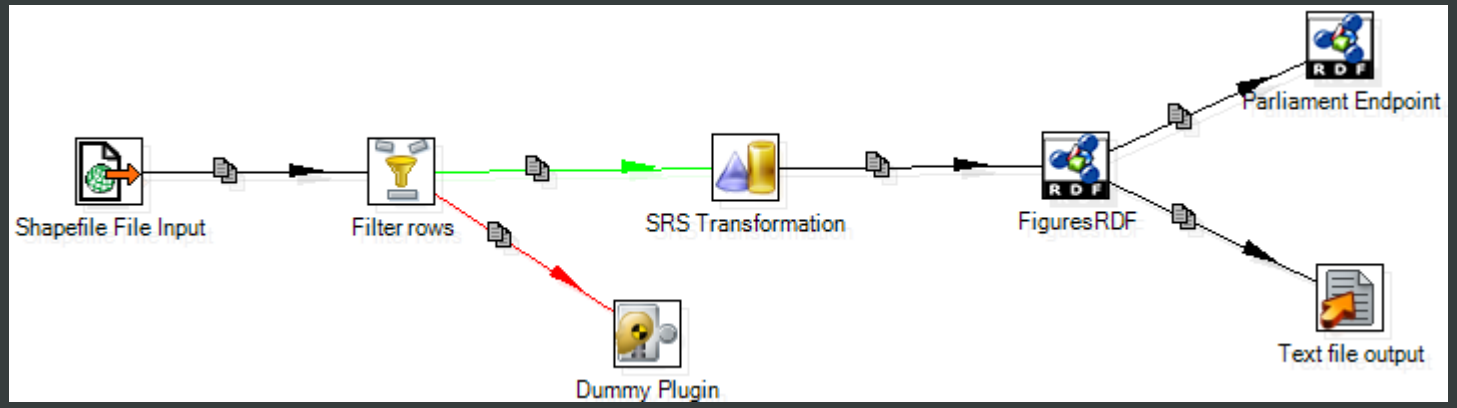
GeoKettle - Generation



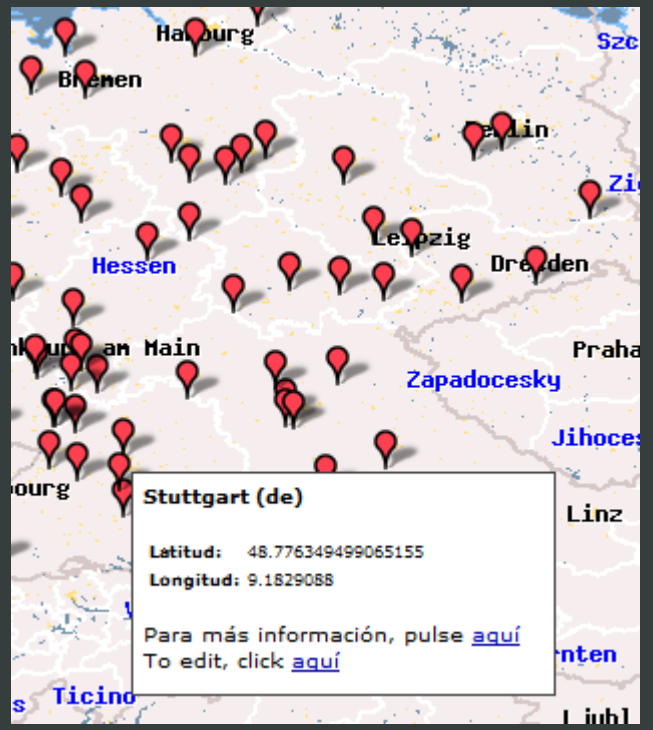
Map4RDF - Visualization

Example of use 3

- Main cities in Germany
- Shapefile from: http://www.mapcruzin.com/free-germany-arcgis-maps-shapefiles.htm
- Using data validation
- Geometry type: Points



GeoKettle Application - Generation



Map4RDF - Visualization

Do you need to develop a new component?

- Documentation available: http://wiki.pentaho.com/display/EAI/Writing+your+own+Pentaho+Data+Integration+Plug-In
- No need to understand all the complex internal GeoKettle functionality.
- No need to modify internal GeoKettle components.
- You only need to build specific tasks.
- And… use existing components!

Future view

Linked Data Warehouses

- Getting information from local and external endpoints (using RDF Input plugin).
- Merging and validating information (through GeoKettle built-in plugins).
- Publishing information in DBs or Endpoints (Parliament Endpoint Plugin).

End-user

- Users would generate their information in RDF easily without programming knowledge.

Integration

- Use of external applications under GeoKettle framework.

The best part

IT'S FREE

http://github.com/marcelocaj/rdf4geokettle

Source code of ***all*** plugins coming soon…