Making Polish geographical names SEO friendly

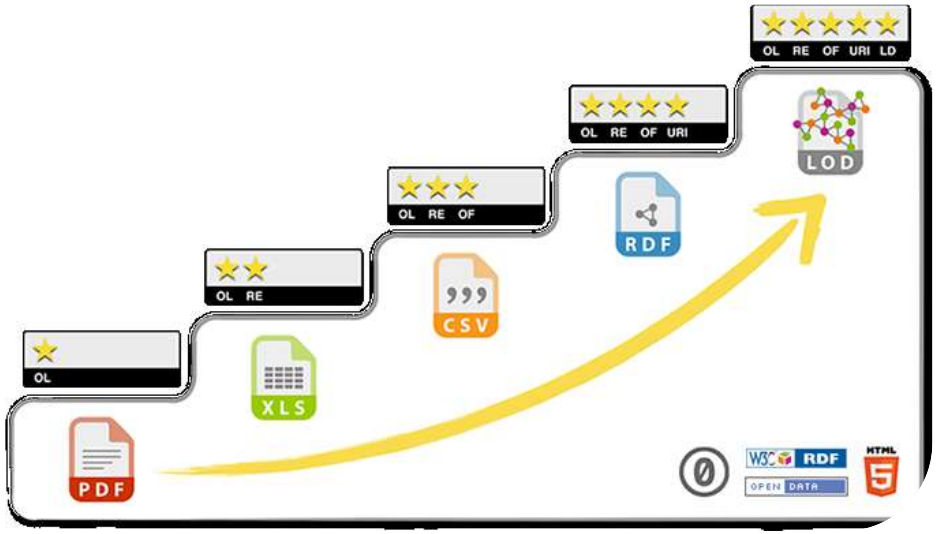
INSPIRE Discovery Workshop 2019

Marcin Grudzień

4th of July 2019

Starting point

- Activities related to publishing spatial data as Linked Open Data (LOD)
- Contract issued
  - create ontologies for State Register of Geographical Names (PRNG) data set
  - transform PRNG to RDF
  - publish created RDF via SPARQL endpoint



Why PRNG?

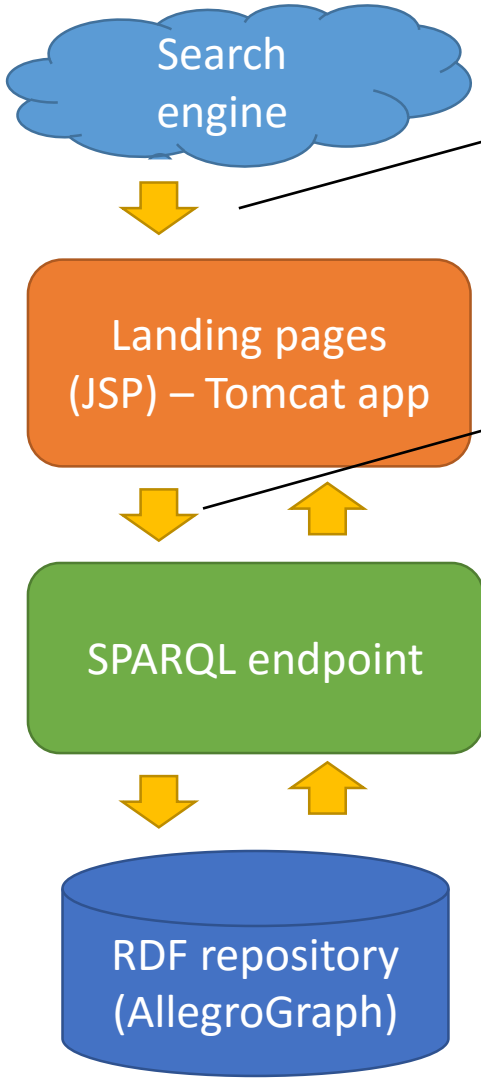
- It is one of the smallest data set maintained by our organisation – contains around 300 000 objects:
  - Build-up areas: cities, towns, villages and their parts
  - Physiographic objects: lakes, rivers, peaks, caves, forests, etc.
  - In general the data set content is similar to INSPIRE DS on Geographical Names
- Point geometry
- Open data
- Referential data set – similar to GeoNames

PRNG as LOD

- At the and of the contract we had:
  - RDF repository (https://dane.gov.pl/dataset/780,panstwowy-rejestr-nazw-geograficznych-prng/resource/1474)
  - SPARQL endpoint (http://semantic.geoportal.gov.pl/)
  - **Unique URI for each individual objects**

- Let’s create landing pages

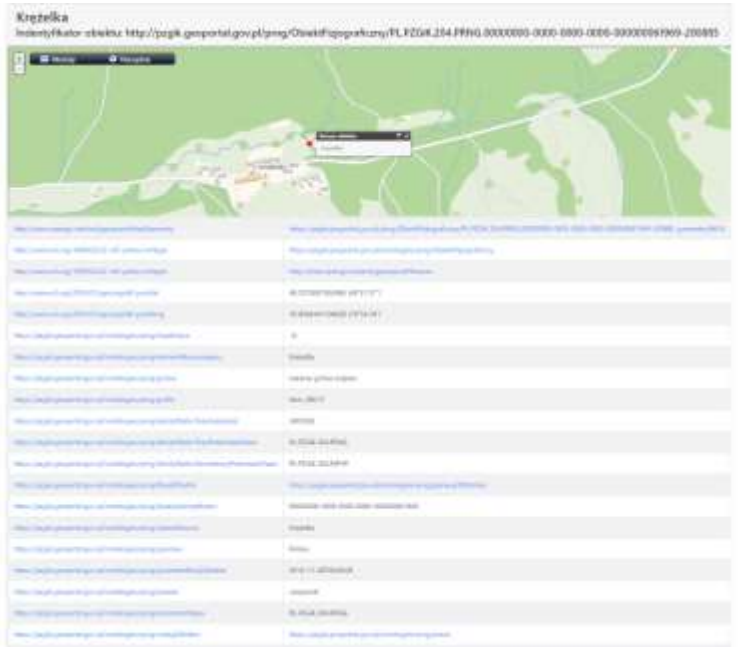
Landing pages



Name of geographical feature
e.g. „Krężelka”

https://pzgik.geoportal.gov.pl/prng/ObiektFizjograficzny/PL.PZGiK.204.PRNG.00000000-0000-0000-0000-000000061969-200885

```
select  ?b ?c
{<https://pzgik.geoportal.gov.pl/prng/ObiektFizjograficzny/PL.PZGiK.204.PRNG.00000000-0000-0000-0000-000000061969-200885> ?b ?c}
```



SEO improvements

- Gradual – stepwise approach
- Sitemaps and Open Search Console
- HTTP to HTTPS
- Adding RDFa
  - Returned by SPARQL endpoint
  - Schema.org

Experience

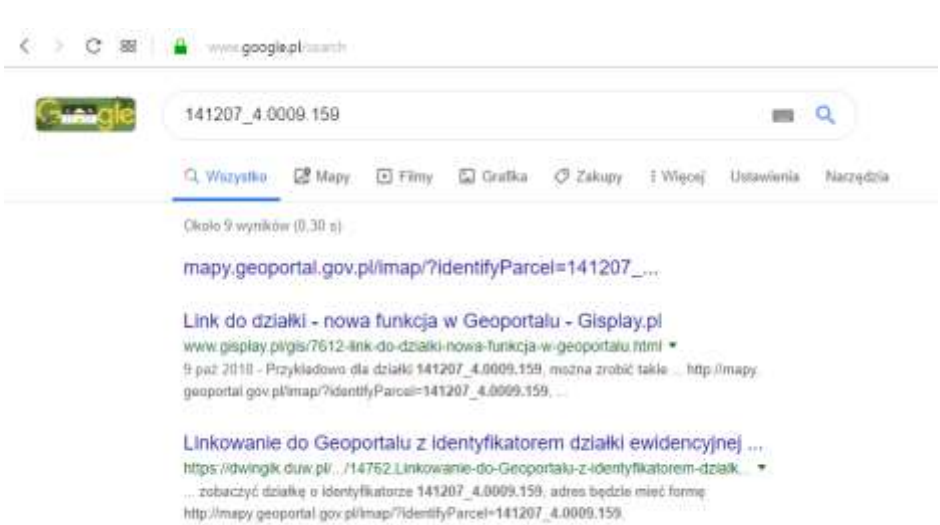
- Google
  - Initially 113 656 objects (landing pages) were indexed out of around 300 000 (33%).
  - At the beginning of this year Google Search Console reported „Crawl anomaly”. As a result almost all previously indexed websites were removed from the Google index.
  - After investigation it turned out that the anomaly was caused by the security policy of the new firewall – the firewall was blocking sessions of web crawlers.
  - At some point only 3 394 landing pages were indexed by Google. Currently, it is 8 273.
  - We plan to attempt to re-index removed landing pages.

Lesson learned

- In general applied approach worked
- Next time we will do it differently
  - Put more „attractive” content (from the perspective of search crawlers) on landing pages from the very beginning
    - Experiment with small subsets of different version of landing pages to check which version is most attractive for search crawlers and later prepare the rest of landing pages using the most effective template
  - Use additional indexing websites
  - Be more careful with network security policy

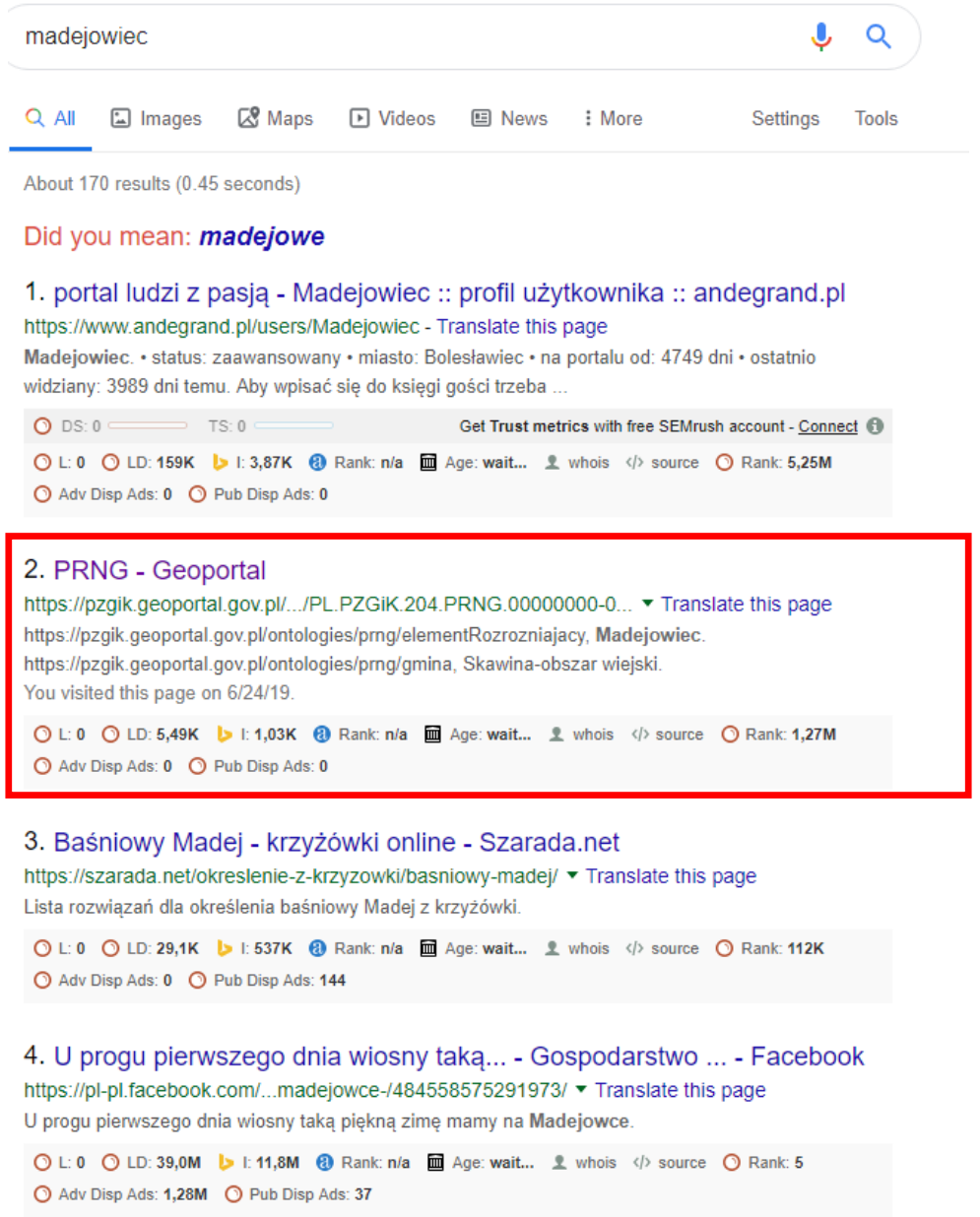
Plans

- Making cadastre data sets SEO friendly
  - Over 35 000 000 cadastral parcels
  - We would like to achieve a situation in which a user would be able to find the landing page of the particular cadastral parcel by providing its unique identifier to a search engine



Live demo

- Madejowiec
- https://pzgik.geoportal.gov.pl/prng/ObiektFizjograficzny/PL.PZGiK.204.PRNG.00000000-0000-0000-0000-000000075208-212526



Thank you

marcin.grudzien@gugik.gov.pl