

# Spatial data and metadata visibility in mainstream search engines: Experiences from Finland

Kai Koistinen, National Land Survey of Finland

# Background

- A lot of open spatial data and services → Have we reached all potential users?
- Linked data pilots → RDF, ontologies, persistent URIs, HTML data cards...
- National recommendation for spatial data URIs
  - URI structure `http://paikkatiedot.fi/{type}/{datasetId}/{localId}`
- Geospatial Platform project
  - <http://www.paikkatietoalusta.fi/en>
  - National platform for nationally harmonized spatial data
  - Intelligent search functionality is one of the services of the platform

# Metadata visibility

- INSPIRE discovery service is implemented with Geonetwork (3.6.0)
  - <http://www.paikkatietohakemisto.fi/geonetwork/srv/eng/catalog.search#/home>
  - Google doesn't index standard Geonetwork metadata pages (or does it now?)
  - Geonetwork offers static html page versions and automatically maintained sitemap
    - We couldn't get Google access and index these pages – because of some network or firewall issue that we couldn't solve
      - Seems that Google has recently indexed at least some dynamic and xml Geonetwork metadata pages
- Finnish open data portal is implemented with CKAN
  - <https://www.avoindata.fi/en>
  - Open spatial data and service metadata is harvested from INSPIRE service
  - Google indexes records from CKAN

# Metadata visibility in Geospatial Platform

- Intelligent spatial data metadata search is a functionality of the platform
  - Metadata is harvested via CSW from national discovery service (Geonetwork 3.2) to Elasticsearch index
  - Uses ontology to get better results for the user
  - Metadata is presented to users with a single html page where content is dynamically fetched from CSW
  - Sitemap for search engines
- JSON-LD schema.org description inserted dynamically to every page's HEAD part (using React Helmet)
  - name
  - description
  - url
  - keywords
  - creator
  - includedInDataCatalog
  - spatialCoverage
- > 2000 indexed metadata records (based on Google search console)
- Schema.org content doesn't seem to be indexed so far

# Data visibility

- Linked Data piloting has been done for example on themes GN and BU
  - Html data card for spatial objects concept has evolved on these pilots
  - Data card implementations on top of
    - WFS 2.0
    - SPARQL
    - WFS 3.0
      - <http://paikkatiedot.fi/so/1000772/12002520>
- Google has indexed some html cards from all of these services
  - No sitemap for any of the services
  - Index content is not very "intelligent" even for RDF data

# Conclusions

- Google and maybe some other search engines can index our metadata and html data cards
- So far the benefits are minimal but we have only taken the first steps
  - You can find pages and data cards by typing correct words or uuid but it's not very "intelligent" yet