



Using search engines to query spatial data

Clemens Portele, interactive instruments

Vision for sharing spatial data

- Features that are of interest to the general public (e.g., buildings, parcels, addresses, protected sites) should be discoverable using search engines
- That is, search engines can be used to query spatial data in a way that complements access APIs like OGC API Features
- Use cases:
 - User wants get relevant information about a/my place and its surrounding
 - Initiate eGovernment workflows related to the place (request a permit, report an issue, etc.)
- Problems / limitations:
 - Spatial data often designed for expert users and for use in mapping
 - Few links to related information
 - No links to actions
 - Need to improve usability
 - Search engines not reliably indexing datasets with millions of features
 - Search engines not using schema:Place information from harvested webpages in rich snippets

Questions for discussion

- Is this the right thing to do? For which data and use cases?
- What needs to be included in the web page to become representative for the real-world thing and "useful"? What information and media should be included? Which links to other information or actions?
- How can we improve the indexing of pages for potentially millions of features? We have looked into sitemaps, schema.org, HTML meta tags (e.g. to identify canonical URIs for each feature). What have others done?

Consolidated summary of the break-out discussions (1/3)

General aspects

- Providing web pages for features should not be driven by search engines in the first place, but feature pages should be primarily for human users
- The targeted human users have diverse needs; they may be data scientists or developers (used to work with datasets and/or APIs, but may have no or limited domain or GIS knowledge), domain experts (understand the data, but may not have the tools to use the data), citizens (are not interested in the dataset, just individual things, will have no domain or GIS knowledge and no access to related tools), etc.
- Web pages for features (i.e. HTML representations of features) can be automatically created by tools consistent with the current JSON/XML/etc. content, but such pages may be of limited use for the target user group
- Identify your target user group(s) and design your data offering accordingly – not only for the HTML representation, this could impact other representations as well
- Identify how the availability of web pages helps with processes of the data publisher or the targeted users – how do the pages improve the user journey?
- Concerns about the usefulness and costs for datasets with many features
- Continue to work on demonstrators to explore, experiment and get feedback
- Publishing feature data with persistent URIs, easy to understand information, directly accessible to the target users can lead to more linking between data, more centralization and less copying
- Potential INSPIRE use cases mentioned:
 - Environmental Reporting Use Case (see parallel break-out by Marc Olijslagers)
 - Look at challenges of the Helsinki Inspire Event

Consolidated summary of the break-out discussions (2/3)

Web pages for features / real-world things

- Web pages should provide information, not just show data, and in general be richer than the typical feature data in most datasets today
- Improvement ideas:
 - Describe the real-world thing in the terms that the targeted users understand / would use, which will often differ from the language that domain experts use
 - Translate codes etc. to text that can be understood by the target user group
 - Embed and/or link relevant related information that is of interest to the target user group (e.g. in a real estate context for citizens and app developers and cadastral parcels as features this could be real estate pricing, statistical information, buildings, points of interest in the proximity, etc.) – this may require cooperation with partners
 - Link to relevant actions that is of interest to the target user group (e.g. receiving a land register report for a parcel, applying for a permit related to a building or parcel, etc.) – public sector datasets support government processes, consider how sharing datasets on the web can help with the digital transformation of such processes
 - Add schema.org annotations
 - Keep information up-to-date and document changes
 - Provide context: for example, include a link from the feature page to the dataset landing page

Consolidated summary of the break-out discussions (3/3)

Indexing by search engines / catalogues

- Indexing feature data may also be useful domain search engine one like the INSPIRE Geoportal (could enable searches on the dataset contents, not just the metadata)
- Consider to index only features/things that are of interest to the target user group or that are linked by others (use noindex/nofollow/etc. to control crawler behaviour)
- Consider using a separate host/domain to avoid impact/dilution of indexing many feature web pages on the ranking of other web pages of the organisation
- Use Google Search Console, Google Analytics or similar tools to understand how users find and use the pages
- Harvesting and cataloguing may also lead to duplicate copies, always state the canonical URI of a web page (for both feature and dataset/distribution resources)
- In the INSPIRE context: Develop good practices for the relationship between INSPIRE URIs of INSPIRE spatial things and the canonical URIs (or update <http://inspire.ec.europa.eu/ids>)?