interactive instruments

GEOCAT
GOVERNMENT GEOGRAPHIC DATA PUBLISHING

Linked Data Factory

SPATIAL DATA ON THE WEB

The *proxy on SDI* approach part II; the *catalogue*

@pvangenuchten march 2016

SOME THOUGHTS

not based on any scientific research (although there probably is), just common sense

- In Spatial and Open Data we typically use Catalogs to register datasets to make them discoverable and assessable
- A large part of the potential users is not looking for a dataset, but for an answer to a specific question, which may be related to a single record in such a dataset

- That user group will use a search engine (or related service) to start the search for an answer
- If too many effort required to find the answer, the user will return to the search engine

- So, how can we feed and educate the search engine to give proper answers, so our open data is optimally used
- Or… can we learn from the search engine which of our open data is most efficient in answering questions?

- What is a catalogue from a search engine perspective?

- A website with links to other websites. Main Search Engines currently don't have specific behaviour for schema.org/Datasets, so what is a Dataset from the perspective of a Search engine?

- A website with a collection of strings (if it is in a crawlable format) or a collection of Things if annotated with schema.org

- Things should link to things, that's how context and knowledge arise

- Links are facilitated by linking to base registries (dbpedia, BAG/BRT/BRP, Google Knowledge graph, Facebook) and using common ontologies (schema.org)

- Each Thing should have a unique persistent resolvable identifier

- In recent years we've invested in setting up data infrastructures (INSPIRE, Basis registraties etc) based on ISO, W3C & OGC standards

- Can we make those infrastructures compatible with data on the web, with little effort?

- Can such an approach be a first step on a route to improvement?

MAYBE

But at least we'll learn a lot on the way

- So let's get started!
- Let's update Nationaal Georegister so it can be crawled by search engines
- Let's add a WFS-to-HTTP proxy for each WFS (and Atom) in there
- Let's engage with the search engines

This is what we did in the last three months

Let me share some outcomes/stands…

Search engines do crawl schema.org structured content, but it's a black box how they (are going to) use it

- The search engine world is hardly compatible with the linked open data world.

- If you want to support both of them, the current best practice is to redirect both of them to separate url's with alternative formats/ ontologies.

- Why can't search engines crawl a website like GeoNetwork?

- The answer is in the fancy javascript frameworks that inject ajax into the DOM. Fancy for developers (and users?), but not optimal for search engines, semantic web, link sharing, accessibility, visitor tracking

The metadata and services referenced from that metadata in Nationaalgeoregister need higher quality

The authoritative linked open data web will get architectured disconnected from the linked open data web (will any judge allow a definition from a community effort like wikipedia?)

Thank you

now got to bing.com/toolbox/webmaster, google.com/webmasters, *and* webmaster.yandex.com *and start learning from your spatial open data*