Workshop Programme

Day 1: Wednesday, 3 July 2019

13:30-14:00 Workshop opening

14:00-15:30 Plenary session 1

Timing: 10 mins per presentation + 5 mins discussion

All roads (will) lead to spatial data. The Italian path

Antonio Rotundo (AgID)

>> Abstract

AgID (Agency for Digital Italy) has been working on a national strategy for an overall public sector information infrastructure as knowledge base of all public information resources, in order to make public data available to an ever wider and diversified audience.

One year ago the new release of the National Catalogue for Spatial Data (RNDT – Repertorio Nazionale dei Dati Territoriali) was launched including some new functionalities aiming at further facilitating and improving search, access and use of spatial data in addition to the usual tools for data discovery (catalogue, general search functions, …).
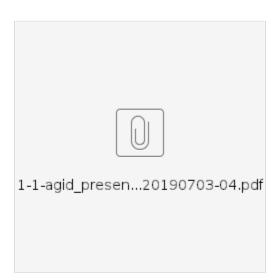
Such tools include the REST interface to query the resources through REST APIs, described using OpenAPI specification.

Furthemore, in order to achieve the interoperable integration between RNDT and the national open data portal, recently AgID released the Italian GeoDCAT-AP API, partially reusing that one developed under ISA programme. The API allows to transform metadata on data sets documented in the national Catalogue, based on Italian metadata profile extending INSPIRE profile, from ISO 19139 records both into DCAT-AP/DCAT-AP_IT used for open data portal and into the GeoDCAT-AP(_IT) extended profiles.

The API works both with CSW requests (GET and POST method) and REST requests and returns metadata in RDF/XML and JSON-LD formats. It can be used by the users to have a different metadata format, but it is also used by the open data portal to retrieve the open spatial data and for an alternative metadata publication.

Next steps planned in this process concern:

- the embedding of Schema.org markup in the web pages and the use of sitemaps. This should be available within the date of the workshop and be presented in that occasion;
- the implementation of additional pre-defined views;
- the finalization and consequent implementation of specific guidelines on the use of URIs for spatial data sets.

1-1-agid_presen...20190703-04.pdf

Making INSPIRE data discoverable and findable through popular search engines: The French experimentation

Abdelfettah Feliachi, Sylvain Grellet, Thierry Vilmus (BRGM)

>> Abstract

Who, today, has never experienced difficulties to find data through INSPIRE specialized search engines like Geoportals or Geocatalogues? Frequently, the general public is even unaware of the existence of such data platforms. When Mr Smith is looking for some piece of information, he will turn to Google or the likes. So, how could our datasets and API endpoints be made discoverable through such search engine?

A possible solution is to expose the national INSPIRE catalogue (the "Géocatalogue") content using Linked Data practices with the Schema.org vocabulary.

An open community process develops Schema.org, founded by Google, Microsoft, Yahoo and Yandex. It is used by popular search engines to index information. In particular, Google offers a set of guidelines to ensure indexing is properly done when adding structured content.

Many possible syntaxes are available to write Linked Data; for this implementation we chose JSON-LD which is versatile: it can be injected into a `<script>` tag in the HTML entries of the catalogue entries, or be declared in static files or over stream.

However, a simple publication of catalogue contents is not enough: a JSON-LD data structure, and above all a Unique Resource Identifier (URI) policy.

In order for this exercise to benefit the entire community we are also pushing our technical implementation to the Geonetwork community (targeting V 3.8.0).

Our suggestions are:

- Use permanent URIs to identify catalogues, datasets and APIs (and hopefully features served);
- Use `schema:Dataset` and `schema:includedInDatacatalog` to link catalogues and datasets;
- Define an URI policy;
- Contribute to open source communities for broader reuse

This system is enhancing the general public experience, allowing Mr Smith to discover, browse, view, download and reuse much more environmental data than before.

It is also increasing visibility and usability of the national catalogue and of the data providers catalogues.

1-2-making_inspi...on_afeliachi.pdf

Towards a Harmonized Strategy on Metadata Catalogues and a Reusable Reference Implementation

Dirk De Baere, **Mathias De Schrijver**, Geraldine Nolf (Informatie Vlaanderen)

>> Abstract

To enable the public and private sector to discover, adopt and reuse government information, administrations publish their data on data portals. The data is accompanied by structural metadata, providing information about the datasets. Governments publish information from different domains, including Geospatial Data, Open Data, Statistical Data, Archival Information, which is causing a wide variety of metadata standards. As these metadata standards often are not interoperable, it is a complex task for government administrations to publish their data in line with the regulations in the different data domains.

To keep it simple and 'once only' for data providers to describe and publish their public sector information on the web, we need to solve the semantic differences in the diverse metadata standards and profiles. This enables data providers to reach a vast audience, and at the same time aligning to all regulations and guidelines applicable to their data domain, describing information once only.

To find out how this could be achieved, a comparative study was initiated by the Flemish Government. It aimed to compare both the Geographic and General Open government data approach towards their respective metadata solutions, with focus on standards, metadata management systems and portals.

The results were crucial with regards to the simplification of the current solutions towards a more costeffective and simplified future implementation. Also, we hope that by closing the metadata gap between the Geographic and the Open government data world, we will clear the road for incorporating other data domains, e.g. statistical data or documents and archival information, APIs.

Next step was to implement the outcome of the study in a pilot. The outcome was proven applicable, and we are glad to announce the upcoming release.

Informatie Vlaanderen and GIM will soon release a DCAT-AP schema plugin for GeoNetwork. The schema plugin is based on an encoding of DCAT-AP v1.1 in RDF/XML syntax. It includes the following features: an enhanced editor, a DCAT-AP RDF harvester, an enhanced RDF endpoint with LDP paging, and an HTML and PDF view of DCAT-AP records. Whereas GeoNetwork was previously mostly used for maintaining metadata about geographical datasets in the ISO 19139 standard, this plugin makes it possible to also maintain records in GeoNetwork using the DCAT-AP standard, which is commonly used for describing open government data. Additionally, the schema plugin makes it possible to bi-directionally convert metadata from one standard into the other. Finally, the harvester and RDF endpoint facilitate a better exchange of metadata between GeoNetwork and other open-data platforms such as CKAN. The plugin will be released under the GPLv2 license by half 2019.

This presentation helps data providers to publish their information not only conform their 'niche' metadata profiles (e.g. ISO and INSPIRE), but as well into the frequently used standards (e.g. DCAT) used in data portals.

1-3-2019-07-03-g...rc_ispra_def.pdf

C3S Climate Data Store: Enhancing the search and discovery of Climate Data and Services

Angel López Alos, **Edward Comyn-Platt** (ECMWF)

>> Abstract

The Climate Data Store (CDS) constitutes the core infrastructure supporting the implementation of the Copernicus Climate Change Service. The system is designed as an open framework able to accommodate changing requirements, evolving needs and distributed data sources. Enhance data discovery and usability was a cornerstone requirement as it was the automatization of the whole integration, publication and dissemination of catalogued resources. In order to achieve this, all the process relies on a centrally managed file configuration system (`*.yaml`) with strict rules for internal organization, naming conventions and metadata content. On top of that automatized deployment (orchestrated by Puppet) and a set of batch processes are triggered to populate the different components with new or updated content. Processes are mostly based on templating (eg. XMLs) and pattern (eg. URLs) replacements implemented using Jinja2 for Python. Among these components are:

- Slugged URLs to harmonized web pages in terms of format and functionality.
- Structured data (JSON-LD format) based on Google Search fields requirements to be embedded on the head of each product web page.
- Compliant metadata records (ISO, INSPIRE) published via CSW (supported by Geonetwork).
- Catalogue Search Engine (supported by Solr);
- DOI registration services (via API calls)

The CDS also provides a set of tools (toolbox), workflows and applications that allow users to perform customized processing, computation, transformation and visualisation on top of catalogued data. Orchestration of workflows relies on the principles described above as also the publication of derived outputs and services.

The aim of this presentation is to demonstrate the orchestration of a complete data publication process, from the file configuration to the final outputs on the web side, summarize experiences and best practices and share for discussion envisaged capabilities considered to be implemented in the coming future.

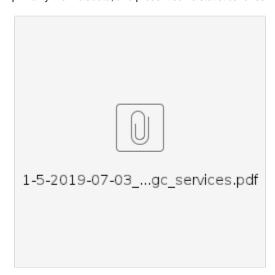CDS is available at: https://cds.climate.copernicus.eu

1-4-c3s_cds_ins...orkshop_ecp.pdf

GeoSeer, A Global Search Engine for OGC Services

Jonathan Moules (GeoSeer)

>> Abstract

GeoSeer is a search engine covering over 1.2 million spatial datasets from over 180,000 public OGC services, including Web Map Services (WMS), Web Feature Services (WFS), Web Coverage Services (WCS), and Web Map Tile Services (WMTS). Launched in early 2018, GeoSeer has global coverage with data from at least 83 countries indexed, including 27 of the 28 EU member states, sourcing data from over 350 data/geo-portals.

GeoSeer has three core components: First, a Scraper to find and collect the GetCapabilities documents; Second, a set of post-processing scripts; and finally, a web-based front-end for making the resultant database searchable. This presentation will largely concentrate on the first component, covering how the scraper works, the problems it encounters, unstable services, outdated links, etc. The presentation will also touch on the post-processing, primarily the metadata, and present some statistics for services, with a focus on the EU and INSPIRE.

1-5-2019-07-03_...gc_services.pdf

15:30-16:00 Coffee break

16:00-17:30 Break-out session 1

Timing: 5 mins per presentation + 1 hour for discussion

Implementation and publication of GeoDCAT-AP metadata by the federal government in Belgium

Benoît Fricheteau (Belgian National Geographic Institute) [presentation not yet available]

>> Abstract

The presentation will briefly explain how the Belgian NGI has been implementing and publishing GeoDCAT AP metadata from the INSPIRE metadata, available on the federal geoportal in cooperation with BOSA responsible for the maintenance of the open data portal of Belgium.

The aim is to set up the federal geocatalogue (holding INSPIRE compliant metadata) as the unique channel to be harvested by the other portals and therefore to avoid the overburden of effort for providing metadata in different formats. The presentation will introduce the preliminary studies done in cooperation with SADL (KU Leuven) on the implementation feasibility, looking at the Belgian case and the relevant outcome and recommendations. Then the presentation will describe the technical processes for generating GeoDCAT-AP metadata records using the JRC's API and the workflow with the main users of GeoDCAT-AP records.

The presentation will conclude on the adaptations needed in the Belgian open data portal in order to proper display geographical metadata and a list of pending issues. We would like to ask people if they are implementing GeoDCAT-AP too, and how they handle maintaining different profiles for the same metadata records.

Making the INSPIRE Geoportal discoverable by search engines

INSPIRE Geoportal Team (European Commission, Joint Research Centre)

>> Abstract

The new version of the INSPIRE Geoportal has been specifically designed to enhance data and service discovery, taking into account the most typical search patterns followed by its users.

The feedback so far was positive, demonstrating a clear improvement compared with the original version of the INSPIRE Geoportal. However, the dataset and service records available from the INSPIRE Geoportal are not yet discoverable from popular search engines, thus reducing their visibility outside the INSPIRE Geoportal itself.

This issue can be addressed by implementing in the geoportal the usual search engine optimisation (SEO) techniques, and those specifically related to dataset discovery.

Most of these techniques can be relatively easily implemented in the current geoportal. For example, annotating the relevant pages with metadata can be achieved by using techniques already experimented in tools used for the implementation of GeoDCAT-AP (see, e.g., the GeoDCAT-AP API and CSW-4-Web), and they can be also automatically complemented with Schema.org annotations (see the DCAT-AP to Schema.org mapping exercise).

Nonetheless, such efforts would be uneffective unless a key requirement is satisfied, namely, the use of "stable enough" URLs for each of the records to be indexed.

Currently, the URL assigned to a given record changes every time it is re-harvested, as no mechanism is in place for recognising whether a record is already in the geoportal or it is a new one. The motivation is that not all the harvested catalogue services include information (as stable identifiers) that would enable the unambiguous identification of a record over time. Consequently, as the priority for the INSPIRE Geoportal is to publish the metadata records available from the INSPIRE infrastructure, stable URLs are a secondary requirement.

Questions for discussion:

1. Should the records available from the INSPIRE Geoportal be discoverable from popular search engines?
2. Should the INSPIRE Geoportal use stable URLs for harvested records? In such a case:
   a. Should the INSPIRE Geoportal use stable URLs only for those records having themselves a stable URL / identifier, and use the current approach for the other ones?
   b. Should the INSPIRE Geoportal use stable URLs for all harvested records? Which techniques can be used for achieving this?
   c. Other options?

2-2-inspire-geoportal.pdf

Role of spatial catalogs in a search engine aware INSPIRE landscape

Paul van Genuchten (GeoCat)

>> Abstract

The WFS3 standard (now called OGC API Features) has been redesigned to follow the recommendations of the spatial data on the web working group. This implies that each dataset exposed via WFS will be crawlable by search engines by default. Potential users will be locating the relevant datasets or features directly on the popular search engines. Which, once adopted as INSPIRE download service, poses new questions: What is the role of the traditional spatial catalog and INSPIRE discovery services? Do governmental organisations need to provide a similar search experience or can it be delegated to the popular platforms? How do end users distinguish authoritative information between others.

2-3-jrc-search-catalogrole.pdf

17:30-18:00 Wrap-up Day 1

Day 2: Thursday, 4 July 2019

8:30-9:00 Registration

9:00-10:30 Plenary session 2

Timing: 10 mins per presentation + 5 mins discussion

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

Kai Koistinen (National Land Survey of Finland)

>> Abstract

For the public administration spatial data a common problem seems to be the lack of awareness of the data in different potential user communities. A lot of data is freely available but it's hard to find and use especially if you are not a spatial data expert. Numerous different catalogue and map applications have been built for the spatial data community but the visibility of the data offered in these applications is still weak in mainstream data searches.

In Finland attempts to improve both spatial data and metadata visibility have been done during the past few years. National recommendation for spatial data URI structure forms a basis for this work. Every data and metadata resource should have a persistent HTTP URI. National recommendation defines a common national structure for the spatial data URIs. National URI redirection service is available for the data providers for better ensuring id persistency.

Finland's INSPIRE discovery service is built with Geonetwork. Support for mainstream search engines is weak in default Geonetwork setup and we haven't been able to improve the visibility of Geonetwork metadata cards. Metadata of Open data and services are harvested to a national CKAN based Open data catalogue and the visibility of these records in mainstream searches is a lot better but the support for spatial standards is very limited. We have a new spatial data platform where we try to improve this situation by using technologies like Elasticsearch, Angular.io, JSON-LD and Schema.org.

3-1-search_engin...on_koistinen.pdf

Making Polish geographical names SEO friendly

Marcin Grudzie (GUGiK)

>> Abstract

Search engine optimisation (SEO) is a process that is often overlooked by public sector data providers resulting in a situation that potential users are unable to find an available data they need. In order to improve in this area Head Office of Geodesy and Cartography (GUGiK) – Polish National Mapping and Cadastral Authority decided to launch a testbed in order to raise the competence and gather experience in this area.

National Geographical Names Register is one of the simplest and smallest data sets maintained by GUGiK. As such it is a good candidate to carry on a testbed. Such testbed was initially launched in 2016 and is gradually upgraded ever since.

The developed testbed currently includes several technologies that support SEO including:

- Linked data,
- Persistent and resolvable URI,
- HTML landing pages,
- Schema.org,
- Open Search Console integration,
- Metadata tags.

The aim of presentation is to demonstrate the technical architecture of the testbed and preliminary SEO results and experiences.

3-2-prng_seo.pdf

Interim findings from the UK Geo Data Discoverability project

Peter Parslow, **Marianne Pope** (Ordnance Survey)

>> Abstract

During the past six months, six UK geo data agencies have collaborated on phase 1 of a Geospatial Commission funded project intended to make it easier to find our data. Many of the initial lessons learned concerned our own different organisational maturity when it came to managing metadata and enabling people to find out data. Each partner published a list of all their datasets, which did result in some interest where it had not even been known that that agency had that data.

A wider group of ten UK stakeholders are now engaged in a stage 2 project, looking to implement the lessons learned, and further trial some of the earlier results. The intention is that this work builds on INSPIRE and the UK national data portal (data.gov.uk); the challenge is that this is not the way most people set about finding data!

You can read more about the project in this January blog post from the phase 1 project leader: "Data Discoverability with Geo6".

This presentation includes two interim results from the phase 1 project which we feel are of wider interest:

- A summary of our user research: how do users actually set out to find geo data? What are their experiences? What can we learn?
- Does embedding Schema.org in your landing page really enhance your search engine rankings?

3-3-data_discoverability.pdf

Experiences with implementing the Spatial Data on the Web Best Practices in NRW

Clemens Portele (interactive instruments)

>> Abstract

The German state North-Rhine Westphalia (NRW) has been exploring how to share datasets of geospatial reference data according to the W3C/OGC Spatial Data on the Web Best Practices and the draft OGC API Features standard (formerly known as OGC WFS 3.0) since 2017. See SDW BP Implementation Report #3 for details.

This includes publishing everything in HTML in addition to XML/GML and JSON/GeoJSON, a landing page for each dataset shared via the API, Schema.org annotations for the dataset, the distribution and each feature, additional links to external resources not included in the dataset itself. As part of this we have also investigated optimising the HTML representation, the meta tags and the Schema.org annotations. To evaluate the effect of changes we have used the Google toolset (Dataset Search, Structured Data Testing Tool, Search Console).

In the talk we will summarize the results and experiences. We hope to receive feedback from others that have made similar investigations.

3-4-190704-expe...ices_in_nrw.pdf

EUMETSAT Earth Observation Data on the Web

Uwe Voges (con·terra)

>> Abstract

Until recently it was more difficult to find and get access to EUMETSAT Earth Observation (EO) Products because the data was only stored in an archive accessible by an EO Portal requiring order processing. For mainstream search engines, it was hard to index the EO products (metadata) as being hidden behind domain specific software interfaces (e.g. OGC CSW) requiring specialized clients.

With the development of an EO On-Line Data Access (OLDA) service the situation changed.

In OLDA, EO Product metadata (e.g. spatial / temporal extent, acquisition details) is distinguished from EO Collection metadata (e.g. application area, satellite and instrument, dissemination channels). Every EO Product is linked to an EO Collection.

In the OLDA Catalogue, the EO products are ingested and provided as Submission Information Packages (SIPs). The SIP files are stored in a cloud-based storage system. A SIP file includes a manifest file, the data object(s) and the EO product metadata. Upon storing a SIP file, the accompanied EOP metadata is indexed by ElasticSearch. Collection metadata is managed separately.

The OLDA EO Data Access API provides a set of modern service interfaces (defined in OpenAPI). The REST-EO API is a simple to use REST-API for browsing EO products by humans (HTML) or by machine clients (GeoJSON). At its basic path it provides the available EO collections. Starting from a collection a client can browse through the assigned EO products based on subsets (defined by a set of axes, e.g. temporal and spatial).

The API is aligned with W3C's Spatial Data on the Web Best Practices and follows principles as known from API's like Amazon's S3.

For indexing mainstream search engines the HTML representations of the collections and specific subsets can be used (registered e.g. via sitemaps).

3-5-eumetsat_e..._uwe_voges.pdf

10:30-11:00 Coffee break

11:00-12:30 Break-out session 2

Timing: 5 mins per presentation + 1 hour for discussion

Which dataset ontology to use in which scenarios

Paul van Genuchten (GeoCat), **Linda van den Brink** (Geonovum)
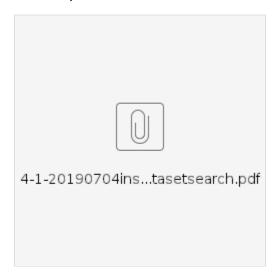
>> Abstract

Quite a number of ontologies are used to describe dataset provenance on the web, like ISO 19115, DCAT, PROV, Dublin Core and Schema.org/Dataset. Software engaged in ingestion or production of these 'metadata records' will want to support as many of these ontologies as possible and provide the records in an ontology relevant to the end user.

These days various software products (LDProxy, GeoNetwork, …) use a convention to return content using a specific ontology at a single URI based on the requested encoding of the output.

- GeoDCAT-ap in `application/rdf+xml` targetting the open and linked data domain
- Schema.org in `text/html` (+microdata) targeting the search engines
- ISO 19139 in `application/xml` targeting the CSW domain

The reasoning behind this is that we try to provide the most relevant ontology to the most likely usertype of the encoding.

We presented this approach at the final conference of the Geonovum Geo4Web testbed. Some suggestions for improvement from the discussion were 1) to introduce an additional content negotiation header to negotiate response ontology and 2) the option to annotate a single record using concepts from multiple relevant ontologies. We would like to continue this discussion with you at the workshop. The discussion is also relevant in the scope of feature discoverability and feature access. WFS3 for example introduces alternative encodings besides XML, such as HTML and JSON. These encodings facilitate the introduction of alternative ontologies, complementing the current UML oriented INSPIRE data models, based on OWL/SHACL, such as Schema.org. Introducing a mechanism to expose a dataset using multiple ontologies facilitates INSPIRE data model conformance as well as search engine discoverability.

4-1-20190704ins...tasetsearch.pdf

Using search engines to query spatial data

Clemens Portele (interactive instruments)

>> Abstract

One of the visions for sharing spatial data in the German state North-Rhine Westphalia (NRW) is that features that are of interest to the general public (e. g., buildings, parcels, addresses, protected sites) can be found using search engines. That is, use search engines to query spatial data in a way that complements access APIs like OGC API Features. We mainly see two related general use cases for this: First, the user wants get relevant information about a/my place and its surrounding. Second, to initiate egovernment workflows related to the place (e.g., request a permit, report an issue).

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?

4-2-190704-using...spatial_data.pdf

Is a linked data approach the way forward to streamline the environmental reporting processes?

Marc Olijslagers (KU Leuven, SADL)

>> Abstract

As part of the project of Reportnet modernisation, EEA is aiming to streamline the environmental reporting processes. In this context KU Leuven explored the options to use the geospatial information covered by INSPIRE in environmental reporting in order to avoid double reporting and resulting possible data inconsistencies. In particular, the geometry of Natura 2000 sites is also reported in the INSPIRE Protected Sites theme. The study tested the possibilities to reference, find and download specific spatial objects required for Natura 2000 through the INSPIRE infrastructure.

To find a specific spatial object first the correct dataset must be identified, followed by locating the specific object in that dataset. The study exposed issues in both steps. With the metadata currently available in the INSPIRE geoportal it is trivial to identify the correct (authorized) dataset(s). Within the datasets there is not always common identifier linking the Natura 2000 reporting information to the correct INSPIRE Protected Site. The absence of a common identifier is often the result of the actual data flow within a Member State, with different administrations for Natura 2000 and INSPIRE Protected Sites, and only one directional communication.

Is the use of a linked data approach the way to avoid these problems?

When a reporting obligation is covered by combining data from different origins (Natura 2000 reporting and INSPIRE reporting), can the use of linked data guarantee reporting synchronisation between the 2 datasets? On the other hand, the current reporting dataflow is slow. The final uniform European dataset reflects a situation often already 1 to 2 years old at the moment of the data release. Can the use of linked data give access to the present version of a protected site at any moment.

Finally, offers the use of a linked data approach offer additional benefits, or introduces new issues that need to be addressed?

4-3-linkeddata_wo...c_olijslagers.pdf

12:30-13:30 Lunch

13:30-15:00 Plenary session 3

Timing: 10 mins per presentation + 5 mins discussion

GeoNetwork as a facilitator of Search Engine Discoverability of ISO 19115 records

Paul van Genuchten (GeoCat)

>> Abstract

These days GeoNetwork records are still hardly harvested by search engines. GeoNetwork typically presents ISO 19115 data in a HTML format, potentially ideal for search engine ingestion. However GeoNetwork by design has some challenges that prevent content to be easily ingested. In this presentation I'll present some optimizations that have been introduced in recent GeoNetwork versions related to search engine ingestion and some experiences with search engine optimization in GeoNetwork from the GeoNovum testbed Geo4Web.

Optimisations focus on the use of Search Engine Console monitoring, use of a sitemap, URI strategy, Schema.org annotations and indicating which parts of the catalog not to crawl.

5-1-jrc-search-geonetwork.pdf

Building Geo-Spatial Data Portal for optimal User experience with CKAN Core-bundle and OPENLAYER Library

Ethelbert Obinna, Richard Figura (CISS TDI)

>> Abstract

Recent trends in digitalization and its applicable use-cases in Industry 4.0, Mobility 4.0 and Smart City technologies raise new challenges on the provision and availability of map-related data and spatial information. Developers and engineers are challenged to make data available in a secure, seamless, and intuitive way. This requires appropriate tools and workflows that gives the best possible user experience.

Solutions like the Kerbal Archive Network (CKAN) provides an essential base for managing metadata. Combined with other open-source tools, such as Open Layer map libraries, it is possible to develop and provide an open data portal – with a focus on spatial information.

Our research project, delves into the CKAN core-bundle with a close integration with our in-house ecommerce platform for spatial data provisioning and management based on the Contao framework.

Amongst others, one main focus for us is to ensure that the geospatial landing page of the portal can - be easily scanned and read by the user, have a fast performance profile, and have the ability to take the first-user-interaction with very minimal effort.

To do this, we ensure a high focus on not just SEO principles, but also on web accessibility principles, and progressive web apps guidelines – both from the device-platform perspective and also from the users' perspectives. With the integration of Apache Solr's search server, we are designing an auto-controlled schema and search-key indexing for a faster response to user queries.

We would love to share our experience with the community on our design-thinking methodologies, our dataset simplification and annotation processes, and our search-key indexing schema.

5-2-ethelbert_o...rkshop_2019.pdf

Searching SELFIE

Katharina Schleidt (DataCove), **Abdelfettah Feliachi**, Sylvain Grellet (BRGM), Nuno Oliveira, Simone Giannecchini, Andrea Aime (Geosolutions)
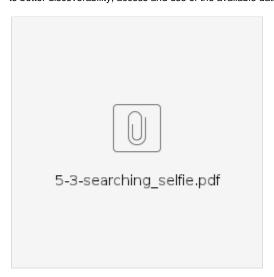
>> Abstract

The initial OGC Environmental Linked Features Interoperability Experiment (ELFIE) sought sustainable and automatable solutions to link multi-disciplinary, multi-organization environmental data without the requirement to transfer custody or burden of maintenance of data. It builds on W3C best practices, providing guidance and a common approach on utilizing JSON-LD for encoding environmental feature data and observational data, as well as semantically defined interlinkages based on Schema.org and other relevant vocabularies.

Using these technologies, it bridged the divide between the great flexibility and power of OGC services and the more focused and specific technologies that drive modern web development.

Building on the content-focused outcomes of the first ELFIE, the Second ELFIE (SELFIE) is designing and vetting Web-resource model and network behavior for cross-domain linked feature data that compliments and uses WFS3 as a building block. This aims to answer the question, how do we use linked data in a way that's compatible with W3C best practices and leverages OGC standards? The experiment aims for focused simplicity, representing resources built from potentially complex data for easy use on the Web. While the IE will test a specific resource model and will follow W3C best practices and OGC standards (WFS and SensorThings API mainly), a wide range of participant-provided domain-use cases will be used for testing. Ultimately, this work is intended to satisfy the needs of many use cases and many kinds of features, from disaster response and resilience to environmental health and the built environment.

In this presentation, we will show how the ELFIEs recommend to utilize linked data technology to bring together disparate geographic features/data items from multiple sources anchored by persistent HTTP URIs, and provide semantic annotation via Schema.org relations between the linked bits, thus leading to better discoverability, access and use of the available data resources.

5-3-searching_selfie.pdf

OpenSearch and JSON-LD for enhanced Earth observation data and service discovery

Ingo Simonis (OGC)

>> Abstract

Discovery of Earth observation (EO) data is an art that requires efforts from several sides to achieve satisfying performance for customers looking for specific data. Data needs to be annotated on the provider side, interfaces need to support simple yet powerful filtering and access mechanisms, and users need to speak the language of the data in order to discover what they are looking for. Recent developments now try to simplify in particular the situation for the data explorer, the user looking for data. These developments use landing pages that provide an entry point into the whole dataset, links from this landing page to the various resources offered by the server, and standardized models for the essential metadata items.

Discovery in the Earth observation domain has featured a different approach in the past. The focus was on maximizing the information provided to the data explorer to allow fine-granular searches in massive datasets using powerful filter languages. This focus led to information models that make sense in their dedicated namespaces and for selected communities only. Several specifications have been developed in this context by ISO and OGC, such as e.g. ISO 19115 Geographic Information: Metadata with its XML serializations and extensions defined in ISO 19139 and ISO 19139-2 to describe geospatial data, and catalog interface specifications with their filtering languages to register and explore metadata entries.

Though proven to be powerful, these approaches cause several challenges for standard search engines. As search engines need to accommodate a broad range of domains, domain specific solutions are usually not or only rudimentarily supported. For that reason, the geospatial community is now exploring approaches that are based on Web APIs with landing pages that provide all required information to explore all resources offered at a specific interface, feature OpenSearch formats and JSON encodings, and try to use domain-independent metadata models extensively. The goal is to integrate as much domain-independent technology and information models and to enrich only where necessary. The goal is to establish a similar user experience for Earth observation data discovery as for website discovery.

This presentation describes the results of OGC's latest discovery experiment: "Earth Observation Process and Application Discovery". The experiment explores the potential of GeoJSON(-LD) metadata encodings combined with OpenSearch formats to enhance the discovery of Earth observation data, processes, and processing services. The initiative features a holistic approach, focusing not only on data discovery, but extends to questions such as "which application can produce the data I need?", "which data is supported by which application?", "which applications can I chain together, given their data input/output capabilities?" By defining a Web-API landing page and corresponding link structure, the initiative explores the level of search-engine friendliness without losing too much discovery capabilities.

The initiative has started exploring the potential of the metadata encodings defined in OGC document 17-084, the GeoJSON(-LD) metadata encoding for EO collections. OGC 17-084 re-uses GeoJSON and OWS Context properties but is designed towards ease of use. It minimizes redesign of properties by re-using properties from several existing namespaces, including Data Catalog Vocabulary (DCAT), Friends of a Friend (FOAF), Location Core Vocabulary (LOCN), PROV Data Model (PROV), Resource Description Format (RDF), Simple Knowledge Organization System (SKOS), vCard Ontology (vCARD), and several others. This metadata side is complemented with OpenSearch response models as specified in OGC document 17-047, the OpenSearch-EO GeoJSON(-LD) Response Encoding Standard. The goal is to use technologies, encodings, and information models that are supported by current search engines to the maximal extent in order to provide a sophisticated user experience to the data and service customer.

5-4-jrc_discover...2019_simonis.pdf

A WFS3 implementation in Python providing Schema.org annotations

Paul van Genuchten (GeoCat)

>> Abstract

While the standard is being developed, various initiatives started around the new WFS3 standard. One of the initiatives, pyGeoAPI, combines the experience of the GeoPython community with the power of Flask, an existing webservice implementation framework. Various implementations of pyGeoAPI are already available these days.

Last month some efforts were put in adding `schema:Dataset` and `schema:DataCatalog` annotations into pyGeoAPI. Having these annotations available makes the WFS services directly ingestable by search engines as datasets and available for dataset search.

In this presentation I'll be showing some of the insights that the team has collected while developing and show some of the potential and impact.

5-5-jrc-search-pygeoapi.pdf

15:00-15:30 Coffee break

15:30-16:30 Conclusions & next steps