Invitation to tender

Spatial data APIs – testbed description and research topics

Geonovum

date
05-07-2021

version
1.0

Revision history	
Revision date	**notes**
Monday July 5th	Published on Geonovum website
Thursday July 8th	Corrected deadline for questions to Thursday July 22nd (instead of 19th)

Table of contents

GEONOVUM

Chapter 1 - Introduction

This chapter gives the general background to the testbed Geonovum is organizing and specifies its goals and scope.

Invitation to tender

This document gives information about the invitation to tender regarding the following five research topics:

- Research topic #1: **CRS extensions for spatial APIs**
- Research topic #2: **Spatial data APIs Discovery**
- Research topic #3: **Spatial data API clients, ease of implementation**
- Research topic #4: **Generic vs Convenience approach for Spatial data APIs**
- Research topic #5: **Simple/linked data encodings for Spatial data APIs**

For each research topic a budget of € 12.500 excluding 21% VAT is available (see chapter 9).

Background

The intended, primary users of the SDI are experts in the geospatial domain. The OGC standards cover the full range of geospatial use cases – some of which are unavoidably complex. Because of this, it requires significant expertise in geospatial information technology to be able to use the SDI. Seen from outside the geospatial domain, the data behind the OGC services is part of the "Deep Web", because the data is published behind specialized web services and not readily available for the majority of web developers [Taylor and Parsons 2015]. This group of users is increasingly making use of, and creating, geospatial data, and is therefore seen by Geonovum as an important new target group to disseminate geospatial data to, in addition to our existing users of the SDI. This brings the following question: 'In which way can, in an evolutionary way, the current SDI be leveraged and the majority of web developers be reached as well?

Geonovum wants geospatial data to be used. The public sector creates a lot of geospatial data, a lot of which is open data and could be useful to others, who are often not experts on spatial data. In addition, we have observed how geospatial data is becoming more and more important for the web and its importance is still growing. Semantically meaningful information on objects is required. Which is not just buildings, roads or waterways; but also e.g. legislative boundaries, permits and ordinances. In our opinion it's very important to integrate spatial data with other data on the Web. This is one of the reasons why Geonovum originally started the Platform Linked Data Nederland as a pilot a few years ago, why we organized the Spatial data on the Web testbed in 2015/2016, and it is also part of the mission of the Spatial Data on the Web working group (SDWWG) that OGC and W3C have formed together, in which Geonovum is participating. Geonovum is also actively promoting the use of APIs within the public sector organizing and participating in the API knowledge platform (apigov.nl).

Geonovum's efforts - with regards to promoting the use of APIs – is well aligned with more recent OGC developments towards a set of API standards. These new OGC APIs are based on the well-known OGC Web Service (OWS) standards, but are more in tune with modern web principles.

The Spatial Data on the Web working group is dedicated (among other things) to do the following:

- to determine how spatial data can best be integrated with other data on the Web;
- to determine how machines and people can discover that different facts in different datasets relate to the same place, especially when 'place' is expressed in different ways and at different levels of granularity;
- to identify and assess existing methods and tools and then create a set of best practices for their use [W3C, OGC 2015].

All this shows that there are multiple layers of access required for geospatial data; or in other words, multiple groups of potential users. One of these groups is the group of geospatial experts, and another is the group of the data users (e.g. web developer, data journalist, data scientist) with no or little specific geospatial expertise. The first group can work with all OGC standards; the second are probably better served with simpler interfaces in a more webby fashion.

One of the users of geospatial data is the public sector itself. In the Netherlands, this is recognized in the existence of several key registries; central databases of, to name a few, all registered addresses, citizens, and cars, but also topographic objects such as roads, waterways, buildings, and smaller objects such as lamp posts and traffic signs. These key registries allow governmental organizations to reuse data created in its different sections. Most key registries already offer some form of API. For instance the registry for buildings and addresses has shown an enormous increase in use since they introduced their API. The new Environmental Act ('*Omgevingswet*') which will come into effect soon has also provided a great boost to the use of APIs for spatial data in the public sector. The fact that APIs help spread the (re)use of spatial data beyond traditional expert users is clear here. What we want to know is what the key success factors and best practices are to help more organizations achieve these goals.

Goal of the testbed

In the years since Geonovum's previous Spatial data on the Web testbed, major innovations have been taking place in geospatial data dissemination and -standards. Notably, the OGC Web Service standards have been evolving into a set of standardized OGC Web APIs. These could play a key role in making spatial data part of the ecosystem of the Web. We are keen to implement and adopt these new standards in the Netherlands, but before we do so, we need to answer several questions about them.

The fact that APIs help spread the (re)use of spatial data beyond traditional expert users is clear. What we want to know is what the key success factors and best practices are to help more organizations achieve these goals. What use cases are best addressed with API standards from OGC and when is it better to use convenience APIs? What are the hurdles to implementation, can we negate existing hurdles through better discoverability, what makes an API truly easy to implement and use in client applications, what are the benefits of simpler encoding formats?

We want to address these questions based upon typical use cases and user questions we have identified.

Geonovum, in line with its mission, is keen to get the answers; and it seeks to involve the market to do so. The actual questions and issues to be addressed are described in this document, combined into five research topics.

Scope

In this testbed we are looking at implementing two types of APIs, OGC APIs and Convenience APIs. We have a testbed platform already in place that implements OGC APIs.

Convenience APIs

Spatial Data on the Web discusses access to spatial data through APIs in section 12. SDW distinguishes three options to provide Web access:

1. *Bulk-download or streaming of the entire or pre-defined subsets of a dataset*
2. *Generalized spatial data access API*
3. *Bespoke API designed to support a particular type of use*

It formulates a best practice (#12) on offering convenience APIs:

> *If you have a specific type of application in mind for your data, tailor a spatial data access API to meet that goal.*
>
> ***Why***
> *Providing access to spatial data via bulk download or generalized spatial data access APIs may be too complex for application developers with relatively simple requirements, if the spatial data or the API is complex to understand or too large to handle in a Web application.* ***Convenience APIs*** *are tailored to meet a specific goal; enabling a user to engage with complex data structures using (a set of) simple queries, including spatial search.*

So where relevant, convenience APIs should ideally be offered to make things easier for developers to use the API. For example, a convenience API could offer easy access to much used functionality or high-level operations for a dataset. Think of easy address lookups for a dataset with addresses, or finding the closest-by feature for a given location.

Testbed platform for OGC APIs

For the testbed Geonovum facilitates and maintains a platform to publish and host API implementations and applications, the API testbed platform at https://apitestbed.geonovum.nl/ . The platform will be extended during the course of the testbed and Geonovum prefers that this platform is used to publish applications. But it is not mandatory to use it in the testbed. For example if it takes too much time to publish results there.

At the time of writing the platform offers several OGC API Features instances, using different kinds of software (pygeoapi, GeoServer, LDProxy and QGIS server). The platform intends to use open source software, because other parties can then freely replicate parts of the platform.

The platform uses so-called containers (Docker) to publish and maintain API instances. These containers can be deployed to a sandbox, to try things out. The sandbox is not intended to be used by third party applications, machines can be deleted without notice or (accidentally) by others. Containers can also be deployed to a public demo area, which is intended to be (more) stable. This demo area can be used for the testbed, for example to test clients or using APIs. The demo area is more strictly managed by Geonovum staff. Publication of (new) API instances is done via Github (https://github.com/Geonovum/ogc-api-testbed).

There is documentation on the platform and publication of services / APIs at: http://apitestdocs.geonovum.nl/.

The platform and it's APIs / services are publicly available. Access to Github to make changes to the sandbox or the demo area is upon request.

Outcome

The results of the testbed are intended to contribute to expand and innovate the Dutch public sector Spatial Data Infrastructure in a direction that takes into account the possibilities in the market today, so that the data will be accessible now and in the future, to both geospatial experts and non-geospatial experts, and for future use in ways we cannot imagine today.

The results of the five research topics of this testbed will be used in the Netherlands to give input to the API knowledge platform(apigov.nl) and the Dutch geoportal Publieke Dienstverlening op de Kaart (PDOK) as well as several Dutch programs: the program for the new Environmental Act and the Dutch INSPIRE program. The results are also intended to provide input to the OGC API standards. In addition, Geonovum will create a Dutch best practice geospatial APIs based on the outcome of the testbed. All insights combined will lead to an implementation strategy for OGC API standards in the Netherlands. This strategy will provide answers on questions on how, when and why parties should start using OGC API standards, both from provider and client point of view. The implementation strategy will be updated over time and thus be responsive to relevant developments in terms of standards, implementations and policy. The overarching goal of the implementation strategy is to guide the adoption of the new generation of standards within the Dutch SDI.

This document

After this introduction, chapter 2 explains the tender procedure. Chapter 3 provides a detailed description of available use cases. Chapter 4 introduces the five research topics. Chapters 5 through 9 describe the research topics in detail. Chapter 10 explains the organization of the testbed in more detail. Appendix A gives the metrics by which proposals are judged.

This document is a draft. Based on questions and comments during and after the tender period we will update this document to clarify questions and remove errors. A final draft will be made available within one week of the question period ending.

Chapter 2 - How to tender

This chapter gives the information about the procedure of tender response.

Rules and procedure

The submission period for the tender starts on July 5th, 2021 with the publication of the Invitation to Tender on Geonovum's website, www.geonovum.nl.

The tender is open to private and public parties, and to combinations of parties (consortia). In the case of a consortium, there is one party who acts as the contact point and contractor on behalf of the consortium for the tender with Geonovum.

Questions about the tender can only be asked by sending an e-mail to info@geonovum.nl, addressed to Frank Terpstra, coordinator of the testbed. Questions should be submitted by Thursday July 22nd. These questions and our answers are collected in an Information Note. We will organize an informational meeting on Thursday, July 15th at 13:30h. The minutes of this meeting will be part of the Information Note. At the latest, this note is published on the website of Geonovum on Thursday, July 29th.

Your tender must be submitted by sending an e-mail to info@geonovum.nl, addressed to Rob van de Velde, director of Geonovum.

The tender is preferably written in English[1] and must at least contain:

- The research topic or topics you are applying for;
- Motivation for the research topic or – topics you are applying for;
- Plan of approach for each addressed research topic (maximum of four pages per research topic);
- References (including e.g. publications, projects, blogs, code on GitHub) and curriculum vitae for performers of the research, showing enough relevant knowledge and experience;
- An indication of the in-kind investment;
- Statement of agreement with the publication of the research results and deliverables under a CC/by license.

All outcome will be available under http://creativecommons.org/licenses/by/4.0/. Deliverables of the research topics, in the form of published data, vocabularies, demonstrators, prototypes and the like, must remain available for at least six months after completion of the testbed.

All source code is preferably available under a "popular and widely used or with strong communities" open source license as identified by the open source initiative. The use of other (non-opensource) licenses will be considered only if well motivated.

The deadline for submitting a tender is Thursday, September 2nd, 2021.

Geonovum will judge the received tenders in the second week of September, according to the criteria stated in appendix A.

Parties are allowed to tender for more than one research topic. However, a contractor is only awarded one research topic, not several. The reason for this is our wish to gain different insights by different parties.

Geonovum will announce which party is selected for which research topic on Thursday September 9th at the latest. All parties who have submitted a tender will be informed about this via e-mail.

Note that reviewers of this document and Geonovum staff[2] are exempt from bidding.

[1] The alternative is Dutch

[2] Employees and Secondments

Chapter 3 - Use cases

This chapter describes use cases that can be employed within the testbed. These are provided as inspiration, you are free to re-use them but can also provide your own. The first phase of testbed execution will leave room to make detailed plans on interaction between research topics.

Use case #1: Groundwater levels & agricultural water usage

In recent years drought has been a problem in the Netherlands. Due to various reasons groundwater levels have been far below normal levels leading to restrictions in water usage (for instance for sprinkling crops in agriculture). The key registry for the underground has started adding detailed data on groundwater levels in the Netherlands to its dataset this year. This data can provide up to hourly variations and can be updated daily. Based on this data it should be possible to provide useful APIs upon which farmers can base their decisions of whether they should turn on their sprinklers that day. If the ground water level is above a certain threshold using the sprinklers is acceptable, when it is too low it is not.

The dataset can be viewed in a viewer here: https://www.broloket.nl/ondergrondgegevens
The groundwater levels are available in this viewer by filtering groundwatermonitoring wells (GMW) on the availability of Groundwatermonitoringnets (GMN) with the monitoring purpose of groundwaterlevels(grondwaterstandsonderzoek). At the time of writing there are two active nets. One next to Assen(kloosterveen) and in one "waterschap de dommel" near de Logt.

The raw datasets is available here: https://service.pdok.nl/bzk/brogldvolledigeset/atom/v1_0/index.xml

Documentation on the XML format is available here: https://bro-productomgeving.nl/bpo/latest/grondwatermonitoring/grondwaterstandonderzoek-gld/gld-berichtencatalogus-uitgiftewebservice

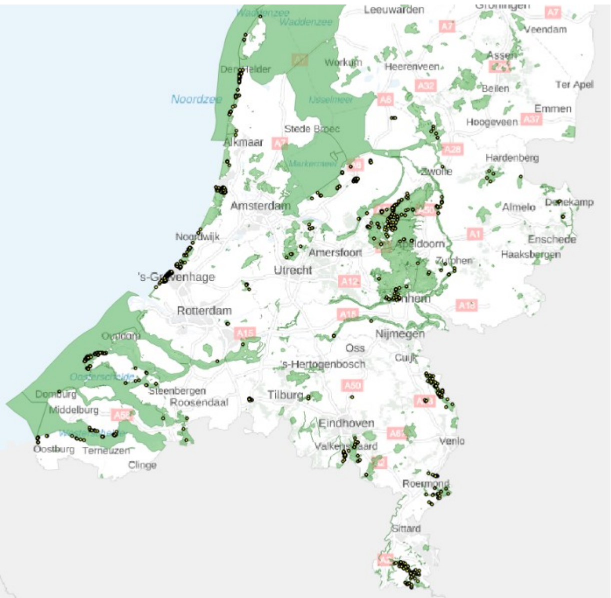
It is based on and compatible with the OGC standard waterML

Use case #2 BGT picnic tables and benches

Inspired upon this geoforum question: https://geoforum.nl/t/zichtbaarheid-banken-en-tafels-aanbod-in-natuurparken-en-bossen/5217

A student is designing a product for blind people and wants to find data about the location of picnic tables and benches in parks and nature.

Dataset: BGT. Data about picnic tables and benches are not available for the entire country, but there are many places where they can be found.



The dataset can be downloaded via https://www.pdok.nl/downloads/-/article/basisregistratie-grootschalige-topografie-bgt-

Documentation on the standards used is here: https://www.geonovum.nl/geo-standaarden/bgt-imgeo

Picnic tables and benches can be found as featuretype "Straatmeubilair" of type "bank" or "picknicktafel". They are not available country-wide, but are available at least in the following regions: …

Use case #3 INSPIRE: nitrogen and Natura2000

The deposition of nitrogen close to or in Natura2000 areas can be harmful and is regulated (see in Dutch, for more information and https://www.aanpakstikstof.nl/ for a broader context). Activities deposing nitrogen need additional checks and permits. Knowing where these Natura2000 areas are and using up-to-date Natura 2000 data in applications and analyses (e.g. for feasibility of new activities or permits) is therefore valuable. Where do these extra regulations apply? Do they apply to the activity someone foresees? For policy makers: do we need to take into account Natura2000 regulations for plans and policies in a specific area?

In addition to these applications, Europe also requires that the Natura2000 dataset is made available to other countries and (central) European agencies in the INSPIRE network for use cases like reporting and cross-border applications. See for INSPIRE. Publication for INSPIRE is regulated: there are technical specifications on how to publish data and metadata. One of the new technical specifications to make data available for download is based on OGC API Features (see https://github.com/INSPIRE-MIF/gp-ogc-api-features/blob/master/spec/oapif-inspire-download.md for the INSPIRE requirements). Such an API is also useful for the local / Dutch applications mentioned before.

In this use case we would like to serve 2 types of users:

1. Users needing access to the dataset itself, for their own applications and reporting. For example engineers, European agencies or local governments. Because of the European context, they require access to the data by the OGC API Features specification of INSPIRE
2. Small / medium businesses planning activities that depose additional nitrogen, that would like to do a check if the location for their activities is in or close to a Natura 2000 area. For example: given a location, is this in a Natura 2000 area or not? Or: what distance is this location from the closest by Natura 2000 area.

Note that this use case oversimplifies the processes around nitrogen deposition and Natura 2000 and that some systems already implement similar use cases (like for AERIUS).

Data that could be used:

https://www.nationaalgeoregister.nl/geonetwork/srv/dut/catalog.search#/metadata/280ed37a-b8d2-4ac5-8567-04d84fad3a41

https://www.nationaalgeoregister.nl/geonetwork/srv/dut/catalog.search#/metadata/8829e5dd-c861-4639-a6c8-fdbb6e3440d2

Because this use case concerns INSPIRE data, special attention should be paid to the work that a dedicated INSPIRE working group has done on this topic: (https://github.com/INSPIRE-MIF/gp-ogc-api-features/blob/master/spec/oapif-inspire-download.md)

Chapter 4 - The research topics

This chapter describes the research topics that are part of the testbed. Each topic is described in more detail in the following chapters.

Introduction to the research topics

The research topics, although overlapping in scope, are specifically chosen to address different leading perspectives and goals. Below these are spelled out, for each of the five research topics.

Research topic #1: CRS extensions for spatial APIs
The goal of this research topic is to demonstrate how spatial data can be used in multiple coordinate reference systems within APIs and to investigate the required effort to a.) serve data in multiple coordinate reference systems and b.) request and use data in specific coordinate reference systems.

Research topic #2: Spatial data APIs Discovery
The goal of this research topic is to make a collection of spatial data discoverable (i.e. FAIR Findable) on the web, using OGC API standards, and to report on how this can best be done.

Research topic #3: Spatial data API clients, ease of implementation
The reasoning behind the use of APIs as a means of publishing Geospatial information is that it makes implementation in client software easier. This should not just be in the domain of geospatial experts. They should be accessible by anyone with generic IT skills. In this research topic we want to put this to the test and discover to what extend this is true for OGC APIs, as well as convenience APIs.

Research topic #4: Generic vs Convenience approach for Spatial data APIs
The goal of this research topic is to discover the usefulness and feasibility of the combination of an OGC API Features implementation and a convenience API. In which cases is it useful to have both generic functions of OGC API Features and convenience functions for a specific dataset in one API? And in which cases it is not? Is it easy to combine both, what are difficulties you encounter?

Research topic #5: Simple/linked data encodings for Spatial data APIs
the goals of this research topic are: exploring the applicability of lighter formats (specifically, JSON) for the publication of geospatial data as well as experimenting with the use of semantically enabled data - focusing the use of the JSON-LD spec in combination with geodata in JSON, and the added value of the output.

Chapter 5 - Research topic #1: CRS extensions for spatial APIs

Goal

The goal of this research topic is to demonstrate how spatial data can be used in multiple coordinate reference systems within APIs and to investigate the required effort to a.) serve data in multiple coordinate reference systems and b.) request and use data in specific coordinate reference systems.

Description

The OGC API Features – Part 2: CRS extension defines the interaction patterns related to the usage of multiple CRS. These interaction patters are different, compared to the interaction patterns described by the Dutch API strategy (based on the idea of content negotiation for CRS). The authors of the OGC API Features – Part 2: CRS specification considered this approach, but decided to define alternative interaction patterns for OGC API Features on this matter.

Task

The task of this research topic is to answer a number of research questions, based on technical implementations (“don't tell them, show them”):

1. Demonstrate how one can serve spatial data in both RD and ETRS89 CRS in such a way that the implementation is OGC API Features – Part 2: CRS-compliant.
2. Demonstrate how the effort, needed to serve spatial data in both RD and ETRS in accordance with OGC API Features – Part 2: CRS, relates to the effort, needed to serve the same spatial data in those CRSs in accordance with the Dutch API strategy (using content negotiation for CRS).
3. Demonstrate how easy or difficult it is to adapt an existing API that uses content negotiation for CRS, to a version that follows the OGC API Features – Part 2: CRS specification, or to a version that supports both mechanisms.
4. Demonstrate, by using at least one client, how users can request data in a specific CRS from a.) an API that implements the OGC API Features – Part 2: CRS specification and b.) an API that implements the Dutch API Strategy mechanism with content negotiation for CRS.
5. Demonstrate how your favorite (i.e. useful / user-friendly / original / popular) existing API that supports multiple CRSs, can be converted into a version that is OGC API Features – Part 2: CRS-compliant.

Deliverables

- Technical implementations (working prototypes) for each research question that are publicly available at least in the 6 months after the testbed ends. They are preferably implemented at the Geonovum Testbed Server.
- Document containing:
  - answers to the questions in this task description
  - Lessons learned that are likely to be relevant to a.) providers of API's with spatial data in multiple CRS, b.) users of API's with spatial data in multiple CRS and/or c.) policy makers that have to decide on whether and when OGC API standards should be implemented in practice.

Requirements / standards

OGC API Features – Part 2: CRS
Dutch API Strategy - Design rules Extensions - 11.3 CRS-Negotiation
RD and ETRS89: Geodetic reference frames in the Netherlands
Relation between RD and ETRS89

Relevant use cases

Providing data in multiple CRS serves multiple use cases. The first type of use cases is cross border usage of spatial data. A data user, working on a Dutch use case, is likely to work in RD CRS in order to be able to combine the data with other data sources from the Netherlands. However, another data user that works on a similar use case, but now in a cross border setting (i.e. covering (parts of) the Netherlands and Germany or the Netherlands and Belgium) would work in ETRS89 in order to be able to integrate data sources from both sides of the border without problems. The INSPIRE data sets are harmonized across Europe to serve this type of use cases.

The second type of use cases where users benefit from data in multiple CRS, are use cases where data is combined from themes that work predominantly in ETRS89 (for instance offshore data) and themes that work predominantly in RD (for instance onshore data). A typical example of such a use case is planning an offshore wind park and connecting it to the electrical grid. In a way, this is another type of cross border data integration, but here the border in question is the one between land and sea.

If necessary we can provide a datasets with multiple CRSs.

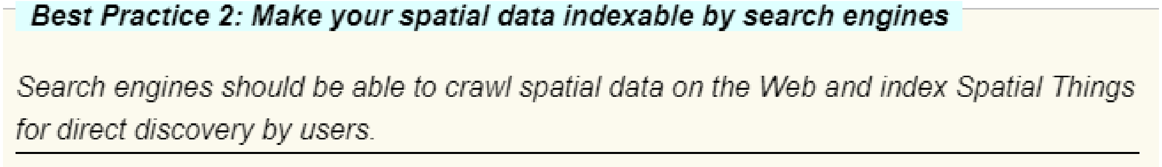
Chapter 6 - Research topic #2: Spatial data APIs Discovery

Goal

The goal of this research topic is to make a collection of spatial data discoverable (i.e. FAIR Findable) on the web, using OGC API standards, and to report on how this can best be done.

Description

In the Spatial Data on the Web Best Practices (SDW-BP), the second Best Practice is formulated as:



Best Practice 2: Make your spatial data indexable by search engines

Search engines should be able to crawl spatial data on the Web and index Spatial Things for direct discovery by users.

Regardless of how fancy your data is, if users cannot find the data, it will never be put to good use. Data portals have traditionally been deployed to solve this problem, but with limited success. Users still need to know the data portal exists. Also, it is often hard to guess which keywords to use, especially for users who are not familiar with the proper jargon. Especially users from outside the geospatial domain, like web developers, data journalists, data scientists etc. are hindered by these obstacles.

How to make spatial data findable on the Web was one of the major research questions in Geonovum's previous Spatial data on the Web testbed (2015-2016). Since that testbed took place, major innovations have been taking place in geospatial data dissemination and -standards. Relevant developments include the publication of the Spatial Data on the Web Best Practices (2017), and the evolution of OGC Web Service standards into a set of standardized OGC Web APIs.

SDW-BP 2 describes how spatial data can be published in such a way, that general Web search engines can index the data, resulting in users being able to find the data using their preferred search engine.

One of the promises of OGC API Features is that this standard improves the discoverability of spatial data because it supports publication of spatial data (i.e. individual features) conform SDW-BP 2. OGC API Records promises to do the same for spatial datasets. By combining these two APIs, it should be possible to dramatically improve the discoverability of spatial data and thus make spatial data more FAIR.

Task

The task in this research topic is **to publish a collection of spatial data and to make it discoverable (i.e. FAIR Findable) on the web, using OGC API standards.**

More specifically, publish a collection of spatial data using OGC API Features and OGC API Records. Implement SDW-BP 2 for both APIs, and implement SDW-BP 13 for OGC API Records. Attempt to get one or more general Web search engines to index the data, monitor the progress and test if, after the data has been crawled, it can indeed be found.

OGC API Records is (at the moment of writing) still a draft standard. As part of this task we ask your reflection on the state and applicability of this standard.

This task requires these more detailed actions:

- Demonstrate a collection of spatial data published using OGC API Features (implement at least part 1 Core) & demonstrate a dataset description for this collection published using OGC API Records (part 1).
    - As part of this task, list existing OGC API Records implementations including an indication of how stable and complete they are (the latter at least for the implementation(s) used in the execution of this task). What is involved in implementing an OGC API Records server instance? What issues did you encounter? Are there any issues that should be solved in the standard before it is finalized?
- Realize a technical collection between an OGC API Records and an OGC API Features. What advantage(s) does this have and what are the issues?
- Demonstrate the crawling metrics of search engines indexing the exposed data and assess how to achieve optimal indexing results. i.e. Describe the Search Engine Optimization (SEO) steps needed to obtain a good discoverability of the data and datasets. Do the OGC API standards in combination with SDW-BP 2 have enough information to succeed? Describe in detail any new insights to improve discoverability.
- Demonstrate the extent to which a user can actually find the data, and dataset, using popular search engines. i.e. how successful were you in getting the data crawled? How high does it end up in search results? Compare these results on the findings of the previous (2015-2016) tested and reflect on the differences in results.
- Demonstrate the application of keywords in day to day language, or other methods to help the user find the data they need without having to be familiar with the jargon used in the dataset itself. E.g., if a user wants data about trees, but does not know there is a dataset “BGT” that contains this, how can she find out? Does adding keywords in normal day to day language (e.g. ‘traffic light’ ipv ‘traffic control installation’, ‘tree’ instead of ‘vegetation object’) improve discoverability? Show how to add these keywords to the API.
- Demonstrate the offering of at least two different data formats and vocabularies in your OGC API Records as well as in your OGC API Features. As part of this task, answer:
    - Which formats/vocabularies would you recommend offering for dataset descriptions? What are the limitations of these formats and with offering a choice of different formats and vocabularies?
    - Which helps discovery more: a serialization of the data and metadata in JSON-LD or HTML or something else? Experiment to see if the serialization makes a difference and record your findings.

Deliverables

- Document containing
    - findings, answers to the questions in this task description
- Published data collection (at least one) via OGC API Features
- Published description of this data collection via OGC API Records
- Technical implementations (working prototypes) for OGC API features and Records with published datasets that are publicly available at least in the 6 months after the testbed ends. They are preferably implemented at the Geonovum Testbed Server.

Requirements / standards

This task requires implementation of:

- Recommendation 5 from the Encodings section of OGC API Features - part 1: Core,
- Requirements class http://docs.ogc.org/DRAFTS/20-004.html#clause-core (http://www.opengis.net/spec/ogcapi-records-1/1.0/req/core) from OGC API Records - part 1: Core,
- Spatial Data on the Web Best practice 2: https://www.w3.org/TR/sdw-bp/#indexable-by-search-engines
- Spatial Data on the Web Best practice 13: https://www.w3.org/TR/sdw-bp/#spatial-info-dataset-metadata

GEONOVUM

Chapter 7 - Research topic #3: Spatial data API clients, ease of implementation

Goal

The reasoning behind the use of APIs as a means of publishing Geospatial information is that it makes implementation in client software easier. This should not just be in the domain of geospatial experts. They should be accessible by anyone with generic IT skills. In this research topic we want to put this to the test and discover to what extend this is true for OGC APIs, as well as convenience APIs.

Description

In this research topic we will look at three type of client implementations for APIs:

- Traditional Geo clients (such as QGIS)
- Browser based clients (based on opensource libraries such as openlayers, leafletjs)
- Standalone clients (applications or apps)

Traditional Geo clients should have out of the box support for OGC APIs, as well as the preceding OGC standards WFS 2.0 combined with GML. one would expect it is harder to integrate them with convenience APIs. Browser bases clients using open source libraries should be able to support both WFS 2.0 and GML as well as OGC API features using Geojson. Convenience APIs should be possible as well as long as they contain Geographic information in a supported encoding format.

For standalone clients one can could imagine either a demo application written completely from scratch or integration with an existing application whose not first and foremost to display geographic information, het Geographical information is just one of the in/outputs it needs.

The reasoning behind comparing the ease of implementation is on the one hand to validate the assumption that APIs are indeed easier to implement but also to discover what hurdles are in the way of widespread client implementations using geographical information provided through APIs. We view the WFS 2.0/GML as a baseline that can be compared to OGC API features as well as convenience APIs.

Task

The task of this research topic is to evaluate the implementation of three types of client accessing Spatial Data APIs. Within the testbed we have available OGC API features APIs as well as one or more convenience APIs. For Each client type try to access Spatial Data on at least on OGC API Features API as well as one convenience API adhering to the API design rules.

Answer the following questions:

- What hurdles exist for nonspatial data specialists accessing each type of API
- Describe good metrics to measure ease of implementation objectively
- Described experience subjectively
- Demonstrate hurdles to implementation explicitly in a demo
- Describe positives of issues found with APIs adhering to API design rules
- High light the most important differences in different client approaches
- What do APIs offer over the existing baseline (WFS 2.0/GML)

Deliverables

- Short report(in English) containing answers to research questions and general findings
- Presentation at final event
- Client implementations available for 6 months after testbed ends
- Published opensource source code for clients

GEONOVUM

Requirements / standards

- OGC API features
- API design rules
- WFS 2.0
- GML

Chapter 8 - Research topic #4: Generic vs Convenience approach for Spatial data APIs

Goal

The goal of this research topic is to discover the usefulness and feasibility of the combination of an OGC API Features implementation and a convenience API. In which cases is it useful to have both generic functions of OGC API Features and convenience functions for a specific dataset in one API? And in which cases it is not? Is it easy to combine both, what are difficulties you encounter?

Description

OGC API Features is a standardized specification for a generic use case: retrieval (and optionally modification) of geospatial features. Because the specification is standardized, it is possible to write generic clients and libraries to consume API implementations, even without knowing which API implementation will be used. But some datasets are often used for specific use cases and simplified API calls could make it much easier to use an API. For example, a dataset with addresses is often used to geocode addresses (find coordinates matching a part of an address string). Such convenience API calls are not part of OGC API Features, because they are not generic. They are nevertheless recommendable, as described in Spatial Data on the Web Best Practice 12.

In the mainstream IT most APIs are designed for a specific use case, application and/or datasets. For example: an API to retrieve data for dataset X uses slightly different semantics (like different URL paths) to retrieve data, than an API to retrieve data from dataset Y. What these APIs mostly have in common, is that they are REST APIs and sometimes also use the same constructs or specifications to describe the APIs, like OpenAPI documentation, and similar encodings (JSON). They often also offer powerful operations for the use cases/data the APIs are designed for, for example for commonly used queries to that API.

This research topic deals with the combination of both: what if a generic API and convenience API are combined?

Task

The task of this research topic is to implement an API that supports the OGC API Features specification (at least core) and that offers some convenience functions for the dataset offered. With this implementation, answer questions like:

1. What approach is used and why? For example: did you use generic software that supports OGC API Features and extend that with custom functions? Or did you take an already existing API and added operations to support OGC API Features? And what is the reasoning for this approach?
2. How much effort does it take to implement such an API? What is easy to do, what difficulties do you encounter? For example does it help or is it a burden with respect to: operation names, data organization, encodings of the data and the description of the API?
3. What do (different) users of the API think of the combination? Does it make sense to them? Which operations do they prefer? And why?
4. What lessons did you learn? Can some (generic) tips and tricks be documented?

Deliverables

- A publicly available API supporting OGC API Features and having convenience functions, to demonstrate findings and pros and cons of a combined generic and convenience API
- Documentation containing
  - findings, answers to the questions in this task description
  - Lessons learned
- If possible: a simple client demonstrating using the convenience API functions
- Presentation at final event

- API implementations available for 6 months after testbed ends
- Published opensource source code

Requirements / standards

OGC API Features (at least Core, other parts optional / where relevant)
OpenAPI documentation for the entire API
Possibly relevant: INSPIRE specification for Setting up an INSPIRE Download service based on the OGC API-Features standard (https://github.com/INSPIRE-MIF/gp-ogc-api-features/blob/master/spec/oapif-inspire-download.md) . For example if use case 3 is implemented.

Chapter 9 - Research topic #5: Simple/linked data encodings for Spatial data APIs

Goal

the goals of this research topic are:

- exploring the applicability of lighter formats (specifically, JSON) for the publication of geospatial data.
- experimenting with the use of semantically enabled data - focusing the use of the JSON-LD spec in combination with geodata in JSON, and the added value of the output.

Description

Spatial Data on the Web Best Practice 4 draws attention to the choices that data publishers face when assessing appropriate formats for publishing spatial data on the Web. Although such a choice can be influenced by the use cases that the provider wants to support, some formats seem to particularly suited for the Web environment - JSON is such an example. And while the OGC API specifications do not mandate encodings to be used by an OGC API implementation, the OGC API Features does recommend supporting GeoJSON for feature data (when possible)*. The fact that GeoJSON is commonly used for geospatial data on the web and is supported by many tools and software libraries. This makes it a great candidate for use cases that involve combining geospatial data with data from other sources. However, there are some limitations to the JSON format in question. How these limitations impact one's choice for publishing data in the format is a topic that deserves some more attention.

When publishing data on the Web, it is also important to supply it with clear and unambiguous semantics. This is especially important when publishing/using data that is inter-related with data from other sources. Although there are different ways to describe semantics, JSON-LD allows one to add semantics directly to JSON. The question is whether this is something that can and should be done out-of-the-box.

Task

There are two perspectives that should be taken into account when carrying out this task: that of a data provider, and that of a user (developer). The task involves publishing at least 1 collection of spatial data using OGC API Features, making it available in (Geo)JSON -with the addition of a JSON-LD context - and assessing the added value of the semantically annotated output, in use cases where data integration and accessibility plays a role.

In order to make this assessment, a simple (browser-based) client application using the published OGC API should be created. The goal of the application is to demonstrate how the semantics introduced through JSON-LD is of added value - for example: if it provides easy access, within the application, to semantically unambiguous and useful links. Part of this task is identifying and using existent vocabularies that were designed with this in mind. Therefore, when describing geospatial features participants should make use of the GeoSPARQL vocabulary, if possible.

The following questions should be answered.
From the data provider perspective:

- How suitable is GeoJSON as an output format for the APIs in question?
    - Is offering data according to the GeoJSON specs straightforward, or could there be requirements for publishing the data that are hard to comply to using this format?
    - Are there any data aspects that are missing when making the data available through this format (such as temporal aspects, complex geometry types, etc)?
- How easily can a OGC API be provided with JSON-LD output, what are the challenges in doing so?

    - Are there already vocabularies available that are can be used to this end (this will depend on the use case)? Are these vocabularies findable? Is GeoSPARQL suitable?
    - Are there any issues with regards to the representation of geometries, when complying to the JSON-LD 1.1 spec?

From the user perspective:

- How easily can (Geo)JSON output from a OGC API be combined with GeoJSON from other (convenience) APIs? Are there any issues that deserve attention?
    - Does using JSON-LD output make it easier to combine the data with that of other sources, or do you quickly run into issues?
- Is it straightforward to process and use the LD output within an application?
    - Are the available tools capable of handling geodata?

Deliverables

- Document containing
    - findings, answers to the questions in this task description
    - Lessons learned
- Published data collection via OGC API Features preferably including a valid JSON-LD context available for at least 6 months after the testbed ends.
- (Presentation and source code of) simple browser-based client application demonstrating applicability of JSON-LD

Requirements / standards

OGC API Features
GeoSPARQL
GeoJSON
JSON-LD (an overview of all JSON-LD specifications, notes and best practices can be found on: https://www.w3.org/2018/json-ld-wg/)

Relevant Use Cases

As mentioned in the task description, de use cases that are applicable to this topic are those in which data integration could play a key role. One of the advantages of OGC APIs serving JSON is the ease of use when trying to combine it with data from other sources, since JSON is commonly used by developers and often well understood. GeoJSON in particular is well supported by tools and libraries. Interesting use cases might revolve around data users that need to: combine data from multiple key registries, combine data from a key registry with data from other governmental agencies, or combine governmental data with volunteered geographic information (such as SMGI, obtained through social media APIs). Adding semantics (in this case, through JSON-LD) is important when accessibility is required – this could also be of added value in use cases that involve data repurposing: using the data for a different purpose than it was intended for.

Chapter 10 - Testbed organization

This chapter describes the organization, conditions, finances and planning of the testbed.

Coordination

The coordinator on Geonovum side is Frank Terpstra (Geonovum), with support roles for:

- Thijs Brentjens (Geonovum)
- Linda van den Brink (Geonovum)
- Gabriella Wiersma (Geonovum)
- Friso Penninga (Geonovum)

For every research topic there is a bi-weekly meeting between Geonovum and each contractor, either at the Geonovum office in Amersfoort or online. The agenda items of these meetings are the progress and any issues or technical questions concerning the details of the research topic.

Open testbed sessions

Work on the five research topics will be carried out in parallel. During this time, Geonovum wants the five research topics to inform each other as much as possible. For this reason, Geonovum will organize three sessions (max. 1/2 day each) with the contractors of all five research topics. This will be done for the purpose of aligning and sharing developments and knowledge between the research topics.

These sessions will be public; the contractors will present their intermediate results to each other and an open group. Anyone who is interested can be present at these meetings. This group has the possibility to discuss in an open way the results with the contractors. The insights gained from this will be used as much as possible by the contractors in their further work on the research topics.

In addition to these meetings, Geonovum will organize a larger public session after completion of the testbed, in which the contractors have the opportunity to present their final results.

Planning

Geonovum will announce which party is selected for which research topic on Thursday September 9th at the latest (see chapter 2). The testbeds starts immediately afterwards, as do the bi-weekly meeting between Geonovum and each contractor.

The open testbed sessions with contractors will take place in the beginning of October, November and December.

In October we will focus on synergy between the contractors and the use case they implement. November will focus on realization and in December the focus will be on reporting results.

The deadline for carrying out the research topics is December 31st, 2021.

Hereafter a public, open session will be organized in which all results will be presented by the contractors and Geonovum.

Finance

For each of the five research topics a budget of € 12.500 excluding 21% VAT is available.

These budgets are intended as a contribution towards the research activities of the contractor. The budgets allow each contractor to carry out research and exploratory activities and to develop demonstrators to try things out. The budget is not supposed to cover the entire research activities of the contractors; an in-kind contribution of the contractors is expected.

Appendix A: Metrics

The criteria by which proposals are judged are:

- The quality of the plan of approach
- The quality of the references and CVs
- Affinity with data (publication)
- Impact on existing workflow in terms of quality, cost, etc... and general architecture
- Contributions towards EU and NL govt. information services and flows
- Planned dissemination of the created work; documentation and `how to' documents, publication strategy relative to the community, licenses used.

Proposals are judged with two metrics: a general score against the key goals of this document (75%), and a further 25% for key elements specific to each of the five testbeds.

All proposals will be scored according to the following metrics:

		Weight
Overall	Contribute to the "Goal of the testbed"	10%
25%	Applicability to the "Scope" as defined	10%
	Further open standards and interoperability	5%
Architecture	Re-use of existing Testbed infrastructure	5%
20%	Application of use case(s)	5%
	Re-useability after testbed ends	10%
Proposal	Plan or approach	10%
30%	Conciseness and specificity of the plan	5%
	Portfolio, References and CV	5%
	Dissemination, Licenses, lasting effect of outreach beyond geo community	10%

A further 25% is awarded for each of the testbeds specific key goals:

Specific	1: CRS extensions for spatial APIs	25%
25%	2: Spatial data APIs Discovery	25%
	3: Spatial data API clients, ease of implementation	25%
	4: Generic vs Convenience approach for Spatial data APIs	25%
	5: Simple/linked data encodings for Spatial data APIs	25%