Best Practices for Publishing, Retrieving, and Using Spatial Data on the Web

Linda van den Brink[a], Payam Barnaghi[b], Jeremy Tandy[c], Ghislain Atemezing[d], Rob Atkinson[e], Byron Cochrane[f], Yasmin Fathy[g], Raúl García-Castro[h], Armin Haller[i], Andreas Harth[j], Krzysztof Janowicz[k], Şefki Kolozali[l], Bart van Leeuwen[m], Maxime Lefrançois[n], Josh Lieberman[o], Andrea Perego[p], Danh Le-Phuoc[q], Bill Roberts[r], Kerry Taylor[s], and Raphaël Troncy[t]

[a] *Geonovum, Barchman Wuytierslaan 10, Postbus 508, 3800 AM, Amersfoort, The Netherlands*
[b] *University of Surrey, United Kingdom*
[c] *MetOffice, United Kingdom*
[d] *Eurecom, France*
[e] *Metalinkage, Australia*
[f] *LINZ, Eastern Oregon University, Auckland, New Zealand*
[g] *University of Surrey, United Kingdom*
[h] *Computer Science School at Universidad Politécnica de Madrid (UPM), Spain*
[i] *Australian National University, Australia*
[j] *Institute AIFB, Karlsruhe Institute of Technology, Germany*
[k] *Geography Department of the University of California, Santa Barbara, USA*
[l] *University of Surrey, United Kingdom*
[m] *Netage, Amsterdam, The Netherlands*
[n] *École des Mines de Saint-Étienne, France*
[o] *Center for Geographic Analysis, Harvard University, USA*
[p] *Joint Research Centre, European Commission, Italy*
[q] *Technische Universität Berlin, Germany*
[r] *Swirrl, United Kingdom*
[s] *College of Engineering and Computer Science, Australian National University, Australia*
[t] *Data Science Department, Eurecom, France*

Abstract. Data owners are creating an ever richer set of information resources online, and these are being used for more and more applications. With the rapid growth of connected embedded devices, GPS-enabled mobile devices, and various organizations that publish their location-based data (i.e., weather and traffic services), maps and geographical and spatial information (i.e., GIS and open maps), spatial data on the Web is becoming ubiquitous and voluminous. However, the heterogeneity of the available spatial data, as well as some challenges related to spatial data in particular make it difficult for data users, web applications and services to discover, interpret and use the information in large and distributed web systems. This paper summarizes some of the efforts that have been undertaken in the joint W3C/OGC Working Group on Spatial Data on the Web, in particular the effort to describe the best practices for publishing spatial data on the Web. This paper presents the set of principles that guide the selection of the best practices, describes the set of best practices for publishing, discovery and t yet emerged.

1. Introduction

Spatial data are important. Firstly, because they have become ubiquitous with the explosive growth in positioning technologies attached to mobile vehicles, portable devices, and autonomous systems. Secondly, because they are fundamentally useful for countless convenient consumer services like transport planning

on the one hand, and for solving the biggest global challenges like climate change adaptation on the other [39]. Historically, sourcing, managing and using high quality spatial data has largely been the preserve of military, government and scientific enterprises. These groups have long recognised the importance and value to be obtained by sharing their own specialized data with others to achieve cross-theme interoperability, increased usability and better spatial awareness, but they have struggled to achieve the cross-community uptake they would like. Spatial Data Infrastructures [33], commonly employing the mature representation and access standards of the OGC are now well developed, but have become a part of the "deep web" that is hidden for most web search engines and human information-seekers. Even geospatial experts still do not know where to start looking for what they need, nor how to use it when they find it. The integration of spatial data from different sources offers possibilities to infer and gain new information; however, spatial data on the Web are published in various structures, formats and with different granularities. This makes publishing, discovering, retrieving, and interpreting the spatial data on the Web a challenging task.

By contrast, the linked data web, as a platform of principles, tools, and standards championed by the W3C promises the level of data discoverability and usability that is readily visible in, for example, search engine results for consumer shopping. The principles are based on proven aspects of the Web such as resolvable identifiers, common representation formats, and rich interlinking of independently-published information, but they add explicit vocabulary management and also tooling that targets the huge web developer community. Can these principles be successfully applied to the world of complex spatial data to achieve the desired usability and utility?

There are, already, many good examples of projects and web services that deliver to these goals, such as spatial data publication platforms in the Netherlands[1], Nanaimo City in Canada[2], or the UK Environmental Agency[3], but also popular spatial data collections such as Geonames[4]. However, spatial data custodians struggle to find the *best* way to publish their data in order to optimize the future impact as more data appears, more tools are developed, and the consumer community grows. Similarly, web devel-

opers as data consumers and tool developers as foundation-stonemasons are demanding an expert consensus to guide their product development.

The W3C and OGC standardization bodies jointly convened a large workshop in London in 2014 where these issues were hotly discussed over two days (https://www.w3.org/2014/03/lgd/). As a result of the interest, enthusiasm and challenges identified there, they proceeded to establish a joint working group to develop, amongst other things, a compendium of best practices for Spatial Data on the Web. This paper is the first publication of the best practices identified and endorsed by the working group, and presents the rationale underlying the selection of best practices proposed.

1.1. Background: spatial data, the Web, and semantics

Any data that has a location component can be viewed as spatial data: its spatial nature means certain operations such as proximity and containment functions have a meaning within the spatial domain. A location component is a reference to a place on earth or within some other space (e.g., another planet, a shopping mall, or a person's brain) and can be many things: a physical object with a fixed location, such as a building or canal; an administrative unit, like a municipality or postal code area, or the trajectory of a moving object like a car. The power of geospatial data is in the opportunity to combine and integrate information based on location. Temporal data has similar characteristics. Spatio-temporal data is common, and suggests the potential for a commonality of approach for spatial and temporal aspects of data.

However, in order to use spatial-temporal data in this way, it has to be available and accessible. Common practice is to publish this data using an Spatial Data Infrastructure (SDI). SDIs are based on a service-oriented architecture (SOA). Existing resources are documented using dataset-level metadata, published in catalogs which are the most important discovery and access mechanism [34]. More detailed metadata describing structure and content of datasets, and service requests and payloads, is far less commonly shared, and whilst standards suitable for some aspects of this requirement are emerging, defining a common or best practice remains a challenge.

The Open Geospatial Consortium, founded in 1994, publishes technical standards necessary for SDIs to work in an interoperable way. These stand-

ards are based on the more abstract standards for geospatial information from the ISO Technical Committee 211 on Geographic information/Geomatics.

Different aspects of the SDI are standardised (see table 1).

Table 1: Standardized aspects of SDIs

Aspect	Description	Reference standards
Discoverability	Annotate resources with metadata	ISO 19115 ISO 19119 [26]
Accessibility	Web services for discovering, viewing, downloading, sharing geospatial raster data (coverages) etc.	OGC CSW, OGC WMS, OGC WFS, OGC WCS
Portrayal	Defining rules for displaying spatial data	ISO 19117 [25], OGC SLD-SE, OGC KML
Information modeling	Describing the contents of information resources, including geometry	ISO 19103 [20] ISO 19107 [21] ISO 19109 [22] ISO 19110 [23]
Data exchange	Defining formats for exchanging the data	OGC GML
Spatial reference systems	Specifying the location on earth of geographical information	ISO 19111

The OGC developed the first standards for spatial data web services as of 1998 and has responded to early architectural trends in the Web (e.g., SOA). While SDIs and related standards were developing, so did the World Wide Web. Web standards like HTML, XML and RDF [13] were created in the nineties as well. While the Web started off as mostly documents with hyperlinks, over the years it evolved to much more sophisticated web applications, including mass applications in which geospatial data was used, like Bing Maps, Google Maps, Google Earth and OpenStreetMap. More recently, XML has been replaced in many web applications by its more lightweight cousin JSON, and RDF is on the rise after a slow start.

As a generic language, RDF can be used to publish geographic information. Its strengths include its structural flexibility, particularly suited for rich and varied forms of metadata required for different purposes. However, it has no specific features for encoding geometry, which is central to geographic information. Several vocabularies and extensions have been proposed for this purpose, including a core RDF/OWL vocabulary for geographic information which is part of OGC GeoSPARQL, but also several others like W3C Basic Geo and GeoRSS.

Figure 1 provides a rough illustration of the main areas of overlap between the Spatial, Semantic and the broader "Web" (characterized by large amounts of unstructured as well as structured data in various forms, and ad-hoc approaches to data publishing, driven by Web-centric skills and technologies without explicit support for semantic or spatial aspects). There are some commonalities between these different communities of practice; however, common practices across all three do not go much beyond the defining protocol of the Web - HTTP. Note that XML is a common format (if HTML variants are included) - but is not necessarily used ubiquitously for data exchange.
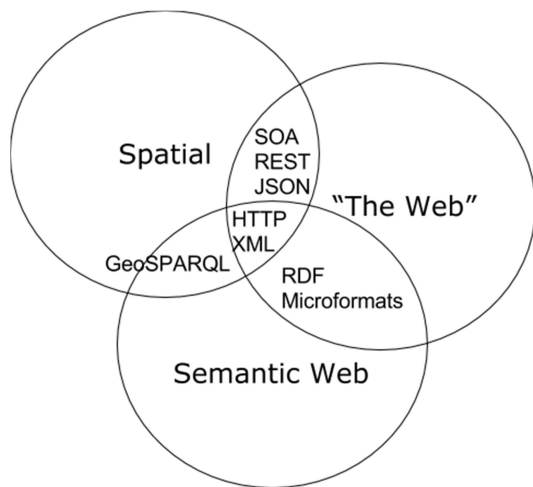
Fig. 1 Commonalities between different communities of practice publishing data on the Web.

1.2. Contributions

The Spatial Data on the Web Working Group (SDWWG) has been committed to determining how spatial data can best be published, discovered, queried and integrated with other data on the Web. This paper is the result of considerable effort on identifying these best practices for publishing and integrating spatial data on the Web. In particular, the paper demonstrates the key requirements for publishing, retrieving and accessing spatial data on the Web to make it more interoperable, accessible, interpretable and understandable by humans and machines, accompanied by know-how and best practices addressing those requirements. In addition, the paper describes areas where best practice is still missing.

1.3. Paper Organization

The rest of the paper is organized as follows: Section 2 explains a set of principles for setting out the scope of the problems that the best practices will address. Section 3 states the key requirements for publishing and sharing spatial data on the Web, presents the related best practices as identified by the working group, and discusses how the best practices address the described requirements. Section 4 discusses several gaps that still exist in current practice. Finally, Section 5 draws conclusions and discusses the future direction.

2. Principles for Describing Best Practices

As explained in the introduction, the aim of the work is to improve the discoverability, interoperability and accessibility of spatial data. The key principle follows from this: that through the adoption of the best practices identified in the document the discoverability and linkability of Geospatial Information published on the web should be improved.

A second principle concerns the intended audience of the spatial data in question. The aim is to deliver benefits to the broadest community of web users possible – not to geospatial data experts only. The term 'user' signifies data user: someone who uses data to build web applications that provide information to end users - web site visitors and app users - in some way. These data users are therefore among the intended audience of not only the spatial data, but also the Best Practice. Application, web site and tool builders addressing the needs of the mass consumer market should find value and guidance in the document. The best practices should be actionable by web application developers who just want to use spatial data without having to become GIS experts; spatial data is just one facet of the information space they work with. Furthermore, the best practices should provide guidance for spatial data custodians. If the best practices can be read as a comprehensive set of guidelines for publishing spatial data on the web, these important stakeholders will be helped a lot.

A third principle is to have a broad focus. The first working draft of the best practices solicited several comments about a perceived "RDF bias". While to develop 5-star linked spatial data was one of the goals at the start, the solutions described in the best practices for discoverability and linkability should also be applicable to other spatial data on the Web. The best practices promote a linked data approach, but without asserting a strong association between linked data and RDF. Linked data requires only that the formats used to publish data support web linking [18]. Furthermore, any ontologies developed within the working group should not be tightly coupled with Upper Ontologies ("compatible with" rather than "dependent upon"); this avoids data publishers having to commit to a given world view, as specified within a particular Upper Ontology, in order for them to use the best practices and any ontologies related to them.

A fourth principle follows from the term 'best practice' very directly: that its contents are taken from practice. The aim is not to reinvent or provide 'best theories;' in other words, the intention of the best practice is not to create new solutions where good solutions already exist or to invent solutions where they do not yet exist. The contents of the best practice should be made up of the best existing practices around publishing spatial data on the Web that can be found. Consequently, the aim is for each of the best practices in the document to be linked to at least one publicly available example(s) of a non-toy dataset that demonstrates the best practice. The best practice document is about the Web, and it should therefore not state something as best practice on the Web if there is nothing attainable on the Web that illustrates the best practice. There should be many (small) examples in the best practice, as examples can be a good way of communicating an idea and are a good way to start an implementation.

Lastly, the best practices should comply with the principles of the W3C Best Practices for Publishing Linked Data and the W3C Data on the Web Best Practices. Where they do not, this will be identified and explained. The Data on the Web Best Practices form a natural counterpart to the work on Spatial data on the Web. The best practices are aligned with the Data on the Web Best Practices (DWBP) in the following ways: a) by using the same best practice template, and b) by referring to the DWBP instead of repeating it. If there is anything specific to add about spatial data, it is incorporated. While the focus of the best practices is on spatial data, they may include recommendations on matters that are not exclusively related to spatial data on the Web, but are considered by the Working Group to be essential considerations in some use cases for publishing and consuming spatial data on the Web, and not covered in enough detail in the Data on the Web Best Practices or other documents.

3. The Key Requirements and Best Practices for Publishing Spatial Data on the Web

A primary purpose of the Spatial Data on the Web Best Practices is to give guidance on how to publish spatial data on the Web, in a way that makes it usable to machines as well as humans, to 1) those already familiar with publishing data on the Web who want to better exploit the spatial aspects of their data; and 2) those who publish spatial data through Spatial Data Infrastructures and want to better integrate that data within the wider Web ecosystem. The Data on the Web Best Practices give general guidelines for publishing data on the Web, but no specific guidelines about spatial data. However, a bit of knowledge about the peculiarities of spatial data can go a long way towards making spatial data more useful to non-expert users.

Spatial data is not special. Most of the architecture used to publish other data on the Web relies on the same principles used for spatial data. But like any other domain, there are some particular issues that it is helpful to be aware of. The better these issues are addressed, the more usable spatial resources published on the Web will be. A primary reason for publishing spatial data on the Web is so that it may be combined with other data – spatial and otherwise. What should we be aware of when publishing spatial data?

3.1. Spatial Reasoning

Tobler's first law of geography states, "everything is related to everything else, but near things are more related than distant things" [40]. Like statistics, in addition to portraying information, spatial is a way of reasoning – to distinguish and divide knowledge. Integration of data based on location is often very useful. For this to be possible either explicit geometries or topological relationships are necessary. Ideally, to enable their widest reuse, geometries should be described having in mind the geospatial, Linked Data and Web communities. This may not be always fea-

sible, but the objective should at least be to describe geometries (also) for web consumption.

One of the key best practices is therefore about providing geometries on the Web in a usable way. A single best way of publishing geometries was not identified; what is 'best' is in this case primarily related to the specific use case and tool support, which determine the geometry format to be used, the Coordinate Reference System (CRS), as well as the level of accuracy, precision, size and dimensionality of geometry data. Note that these aspects are interrelated: for instance, the dimensionality of a geometry determines the set of coordinate reference systems that can be used, as well as the geometry encodings / representations.

The best practice identifies three scenarios in which geometries can be used: specific geospatial applications, linked data applications, and web consumption; and offers guidelines for choosing the right vocabularies from several available ones for describing geometry in each scenario. Currently, there are two reference geometry formats widely used in the geospatial and Web communities, respectively, GML and GeoJSON. GML provides the ability to express any type of geometry, in any coordinate reference system, and up to 3 dimensions (from points to solids) but is typically serialized in XML. GeoJSON supports only one coordinate reference system (CRS84 - i.e., WGS 84 longitude/latitude), and geometries up to 2 dimensions (points, lines, surfaces) but is serialized in JSON, which is often easier for browser-based web applications to process. In the Linked Data community, several specific vocabularies for RDF-based representations of geometries are available; the best practice offers a comparison of the most common geospatial ontologies.

Instead of catering for a single scenario, spatial data publishers should offer multiple geometric representations when possible, balancing the benefit of ease of use against the cost of the additional storage or additional processing if converting on-the-fly. This can be implemented using HTTP content-negotiation; however, this only works for media-type, character set, encoding and language. Consequently, it is not possible to select one representation that conforms to a given "profile" (e.g., data model, complexity level, CRS) from several that all share the same media-type.

Note that publishing geometries on the Web need not always be called for. Although spatial relationships can often be derived mathematically based on geometry, this can be computationally expensive. Topological relationships such as these can be assert-

ed, thereby removing the need to do geometry-based calculations. Exposing such entity-level links to web applications, user-agents and web crawlers allows the relationships between resources to be found without the data user needing to download the entire dataset for local analysis.

3.2. Coordinate Reference Systems and Projections

The key to reasoning and sharing spatial information is the establishment of a common coordinate reference frame in which to position the data. For Earth-based data, this may be done in spherical coordinate values such as latitude, longitude and (optionally) elevation or in a projected Cartesian coordinate space. The latter involves the flattening of a sphere in exchange for making it vastly easier to accurately measure area and distance. Regardless of the Coordinate Reference System (CRS) chosen, a distortion of the data will incur either in relative angles (positions), sizes (areas), or distances. The best practice identifies the World Geodetic System 1984 (WGS84 – EPSG::4326), which provides a good approximation at all locations on the Earth, as the most commonly used CRS for spatial data on the Web, but also explains when EPSG::4326 is not recommended, especially in use cases that require a higher level of accuracy than WGS84 can offer.
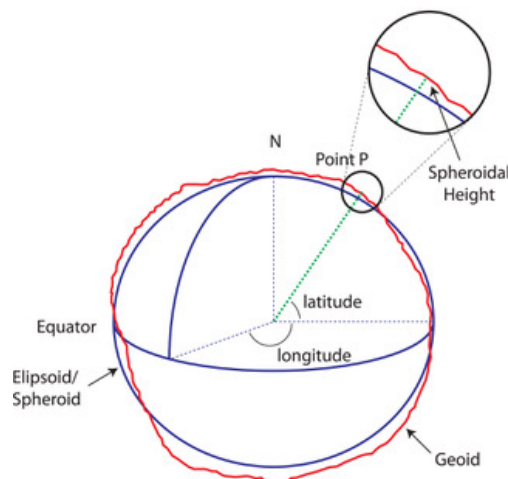

Fig. 2 Elipsoid and spherical coordinates. ("Creative Commons Attribution 3.0 Australia Licence " from ICSM.gov.au Intergovernmental Committee on Surveying and Mapping)

All spatial coordinate frameworks begin with a mathematical model of the object being mapped. For the Earth, (which is not a true spheroid) an ellipsoid

model, most fitting to the area being mapped, is commonly used (figure 2). WGS 84 (EPSG::7030) is an example of a reference ellipsoid. A Geodetic Datum is then placed on top of the reference ellipsoid to allow numerical expression of position. This may include a Vertical Datum, usually an approximation of sea level, to reference height and depth. WGS 84 (EPSG::6326) is an example of a Geodetic Datum based on the WGS 84 (EPSG::7030) Ellipsoid. It uses Latitude and Longitude coordinate values (anchored to the equator and poles) to indicate position.

WGS 84 (EPSG::4326) is an example of a two-dimensional Geodetic Coordinate Reference System and is based on the WGS 84 (EPSG::6326) geodetic datum. It uses Latitude and Longitude in either Decimal Degrees (DD) or Degrees Minutes Seconds (DMS) to specify location. EPSG::4326, EPSG::6326 and EPSG::7030 are all commonly referred to as WGS 84. The latter two are used separately or together in many other CRS's as well. Some of these are also called WGS 84, so use of EPSG codes is recommended for avoiding confusion.

Fig. 3 The website http://thetruesize.com/ is a good tool for comparing sizes of countries at different latitude and shows the distortion which results from flattening a sphere.

Geodetic CRSs, especially EPSG::4326, are useful for collecting information within a common frame. The measurements they use are good for plotting directions but difficult to use when calculating area or distance. Latitude and Longitude are angular measurements that do not convert easily to distance because their size in true units of measure (e.g., me-

ters) varies according to location on the sphere. They become smaller as you near the poles.
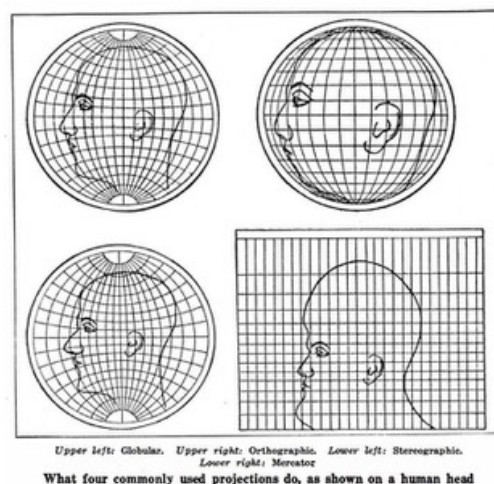

Fig. 4 Projected CRS. (Public domain image from Scientific American circa 1923)

In order to measure size and distance, and to display spatial information on a screen or paper, Projected Coordinate Reference Systems are used. A Projected CRS flattens a sphere to enable it to be portrayed and measured in 2 (or 3) dimensions (figure 4). Flattening a sphere introduces distortions. Imagine flattening the skin of an orange. You can preserve fairly accurately measurements for a portion of the skin but the rest will necessarily be distorted, either through stretching, compression or tears (see figure 3). Projected CRSs are optimised to preserve distance, area or angular relations between spatial things. With Projected CRS measurements are now possible with good accuracy for a chosen region. Distortion grows the further you move from that region. This is one reason why there are so many CRSs. Spatial data should be published in CRSs most common to potential users.

3.3. Spatial Identifiers

Spatial things should be uniquely identified with persistent URIs in order for those using spatial data on the Web to be able to definitively combine and reference these resources: they become part of the Web's information space; contributing to the Web of Data. Uniform Resource Identifiers (URI) are a single global identification system used on the World

Wide Web, similar to telephone numbers in a public switched telephone network. HTTP(S) URIs are a key technology to support Linked Data by offering a generic mechanism to identify entities ('Things') in the world and to allow referring to such entities by others. Anyone can assign identifiers to entities ('Things') in a namespace they own - e.g., hospitals, schools, roads, equipment, etc.

'Spatial Things', such as the catchment area of a river, the boundaries of a building, a city or a continent are examples of such 'Things' on the Web that need to be identified so that it is possible to refer to or make statements about this particular spatial thing.

Spatial Things described or mentioned in a dataset on the Web should be identified using a globally unique URI so that a given Spatial Thing can be unambiguously identified in statements that refer to it. Good identifiers for data on the Web should be dereferenceable/resolvable, which makes it a good idea to use HTTP - or HTTPS - URIs as identifiers. Data publishers need to assign their subject spatial things HTTP URIs from an internet domain name where they have authority over how the web server responds. Typically, this means minting new HTTP URIs. Important aspects of this include authority, persistence, and the difference between information resources and the thing they give information about. The use of a particular internet domain may reinforce the authority of the information served. Designing a URI pattern that is persistent requires careful thought; organization names, for example, are not considered persistent and should be avoided. By its nature, HTTP can only serve information resources such as web pages or JSON documents. Yet a Spatial Thing is a real or conceptual phenomenon. However, although using a single URI to refer to both the Spatial Thing and the page/document that describes the Spatial Thing introduces a URI collision, the best practice contains no requirement to distinguish between the Spatial Thing and the page/document unless an application requires this.

When considering HTTP URIs for objects (e.g., the target of our hyperlinks) it makes sense to "reuse" existing identifiers, i.e., link to URIs available in community-maintained web resources such as DBpedia and GeoNames or public government data, such as a national registers of addresses. Mapping and cadastral authorities maintain datasets that provide geospatial reference data. Re-using well-known identifiers is a good practice because it makes it easy to recognise that data from different sources are related. An example of a Spatial Thing URI is http://sws.geonames.org/2172517/ which identifies the Spatial Thing *Canberra*, while when resolved in a browser returns information such as the name of the Spatial Thing, the centroid location, geographical boundaries, etc.

A Spatial Thing can change in different ways, and sometimes the change is such that it is debatable if it is still the same thing. Therefore it is recommended to version the identifiers for Spatial Things and make statements about the temporal aspect of when the statements about the Spatial Thing are valid. If different versions are provided, it is recommended to provide a canonical versionless URI that resolves to the latest version of the information of the Spatial Thing, as well as date-stamped versions.

When exposing spatial data through standard SDI's (e.g., WFS services), a certain user group is catered for, i.e., users with the expertise and tooling to use these services based on standards from the geospatial domain. To allow more users to benefit from the data it is important to expose the link to the web representation of the features on the Web. The Spatial Data on the Web Best Practice identifies two approaches for doing this while keeping the SDI in place. One is to add an attribute to all spatial things, named 'rdf_seealso', which contains the URI of the web representation of the spatial thing visible on the map. The web representation is created by mapping the data in the SDI dynamically to crawlable resources on the web using the R2RML standard and Linked Data Publication tools. This approach leverages existing SDI's while enriching them with dereferenceable linked data representations of the spatial things. Exposing the data about a spatial thing as linked data makes sure that the attributes themselves will be URI's and thus unambiguous. The overhead of the extra attribute on existing SDI's is neglectable and traditional clients will not be hindered by the extra attribute, but more advanced usage allows unlocking of a wealth of connections behind the traditional spatial data.

The other approach is to create a RESTful API as a wrapper, proxy or a shim layer around WFS services. Content from the WFS service can be provided in this way as linked data, JSON or another web-friendly format. There are examples of this approach of creating a convenience API that works dynamically on top of WFS such as the experimental ldproxy. This is an attractive option for quickly exposing spatial data from existing WFS services on the Web. The approach is to create an intermediate layer by introducing a proxy on top of the WFS (data service) and CSW (metadata service) so the contained resources

are made available. The proxy maps the data and metadata to schema.org according to a provided mapping scheme; assigns URIs to all resources based on a pattern; makes each resource available in HTML, XML, JSON-LD, GML, GeoJSON, and RDF/XML (metadata only); and generates links to data in other datasets using SPARQL queries.

3.4. Discovery of Spatial information

Cataloguing of spatial information has always been difficult, whether the data is digital or not. A road map for Wellington NZ may be filed under NZ, Wellington, Transportation, tourism or a large number of other categories. Spatial data therefore has a greater requirement of metadata. The best practice recommends the inclusion of spatial metadata in dataset metadata. As a minimum, the spatial extent should be specified: the area of the world that the dataset describes. This information is necessary to enable spatial queries within catalog services such as those provided by SDIs and often suffices for initial discovery. However, further levels of description are needed for a user to be able to evaluate the suitability of a dataset for their intended application. This includes at least spatial coverage (continuity, resolution, properties), and representation (for example vector or grid coverage) as well as the coordinate reference system used.

In Spatial Data Infrastructures, the accepted standard for describing metadata is ISO-19115 [24] or profiles thereof. To provide information about the spatial attributes of the dataset on the Web, DCAT [31] is recommended. An experimental application profile of DCAT for geospatial data, GeoDCAT-AP [15], can be applied to more fully express spatial metadata. In addition, several geospatial ontologies, which is already mentioned in section 3.1, allow the description of spatial datasets.

3.5. Scale and quality

The quality of spatial data, as that of any other data, has a big impact on the quality of applications that use such data. This is intensified in the Web scenario, where data could be consumed with unforeseen purposes and data that are useful for a particular use case could come from plenty of sources.

Therefore, having quality information about spatial data on the Web significantly facilitates two main tasks to the consumer of such data. One of them is the selection of data, allowing to focus on data that

satisfy the needs of a concrete use case. For example, spatial data accuracy is important when using them in the application of self-driving cars; guiding an autonomous vehicle to a precise parking spot near a facility that has a time-bounded service and then booking the charging spot requires accurate location data. Another is the reuse of data, i.e., understanding the behavior of data in order to adapt its processing (e.g., by considering data currentness, refresh rate or availability).

A fundamental concept of spatial data is the scale of the representation of the spatial thing. This is important because combining data designed to be used at differing scales often produces misleading results. Scale is often represented as a ratio or shortened to the denominator value of a ratio. A 1:1,000 (or 1,000) scale map is referred to as larger than a 1:1,000,000 (or 1,000,000) scale map. Conversely, a million scale map is said to be a smaller scale than a thousand scale map. Data collected at a small scale is most often more generalized than data collected at a larger scale. Concepts related to scale are resolution (the smallest difference between adjacent positions that can be recorded), accuracy (the amount of uncertainty - how well a coordinate designates a position on the actual Earth's surface) and precision (the reproducibility of a measurement to a certain number of decimal places in coordinate values). Figure 5 explains the difference between accuracy and precision.
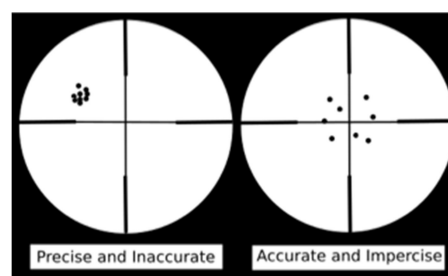

Fig. 5 Accuracy versus precision.

When publishing spatial data on the Web, they should be supplemented with information about the precision and accuracy of such data. Such quality information should at least be available for humans. One of the Spatial Data on the Web Best Practices asserts a link between CRS and data quality, because the accuracy of spatial data depends for a large part on the CRS used, as was explained in a previous section. In order to support automatic machine-

interpretation of quality information, such information should be published following the same principles as any other data published on the Web. The CRS of geometries on the Web should always be made known. For describing other quality aspects, the Data Quality Vocabulary [3], which allows specifying data quality information (such as precision and accuracy), could be used.

Even if the recommendation focuses on precision and accuracy, evidently the same advice can be followed for other relevant spatial quality information.

3.6. Thematic layering and spatial semantics

Spatial data is typically collected in layers. Although this sounds very map-oriented, these layers can be thought of as collections of instances of a class within a spatial and temporal frame. In other words, layers are usually organized semantically. Although the Spatial Data on the Web Best Practice does not address layers directly, it does address spatial semantics. The Data on the Web best practice recommends the use of vocabularies to communicate the semantics, i.e., the meaning of data, and preferably standardized vocabularies.

There are several vocabularies about spatial things available such as Basic Geo vocabulary [5], Geo-SPARQL [35] or schema.org[6]; overviews are provided in the literature [8, 6, 41]. One of the Spatial Data on the Web Best Practices identifies the main vocabularies in which spatial things can be described when the aim is data integration; however, it does not recommend one of them as the best. Currently there is no common practice in the sense of the same spatial vocabulary being used by most spatial data publishers. This depends on many factors; furthermore, describing spatial data multiple times using different vocabularies maximizes the potential for interoperability and lets the consumers choose which is the most useful. Appendix A of the best practices document offers guidance for selecting vocabularies by providing a table comparing the most common spatial data vocabularies.

The most important semantic statement to be made when publishing spatial data - or any data - is to specify the type of a resource. The Basic Geo vocabulary has a class SpatialThing which has a very broad definition. This can be applicable (as a generic concept) to most of the common use-cases. Thematic semantics and general descriptions of spatial things

and their properties should be provided as linked data. They should have URIs that return human and machine readable definitions when resolved.

3.7. Temporal Dimension

It is important to consider the temporal dimension when publishing spatial data. The "where" component of the data is seldom independent of the "when" which may be implicit or explicit to the data. Hence, capturing the temporal component makes spatial data more useful to potential users, since it allows them to verify whether the data suits their needs. For instance, tectonic movements over time can distort the coordinate values of spatial things.

This is included in Best Practice 7 on coordinate reference systems as valuable knowledge when dealing with spatial data for precision applications. Furthermore, it is recommended in Best Practice 13 to include metadata statements about the (most recent) publication date, the frequency of update and the time period for which the dataset is relevant (i.e., temporal extent).

Apart from the need of enhancing spatial data with their temporal context, the temporal dimension also affects the very nature of spatial things, since both spatial things and their attributes can change over time; this is covered in Best Practice 11.

When dealing with changes to a spatial thing, its lifecycle should be taken into account; in particular, how much change is acceptable before a spatial thing can no longer be considered as the same resource (and requires defining a new resource with a new identifier). Creating a new resource will depend on whether domain experts think the fundamental nature of the spatial thing has changed, taking into account its lifecycle (e.g., a historic building replaced by another that has been built on top of it). In this case, the temporal dimension of spatial data can be expressed by providing a series of immutable snapshots that describe the spatial thing at various points in its lifecycle, each snapshot having a persistent URI.

In those cases when the spatial thing itself does not change over time but its attributes do, the description of the spatial thing should be updated to reflect these changes. When a spatial thing has a small number of attributes that are frequently updated (e.g., the GPS-position of a runner or the water level from a stream gauge), the time-series of data values within such attributes of the spatial thing should be represented. This is relevant in relation to recent advances in embedding smart sensors and actuators in physical ob-

---

jects and machines such as vehicles, buildings, and home appliances, which has led to the publishing of large volumes of data that are spatio-temporal [1].

3.8. Size of spatial datasets

Spatial data (and sometimes spatial things) tend to be very large. This can pose difficulties when sharing or consuming spatial data over the Web – particularly in low bandwidth or high latency situations. Accurate (polygon) geometries tend to contain a high number of coordinates. Especially when querying collections of spatial things with geometries over the Web, this results in very large response payloads wasting bandwidth and causing slow response times, while for some very common use cases, like simply displaying some things on a web map, high accuracy is not required. The primary basis for simplification is scale. For example, when searching for a Starbucks on a city scale, an accuracy of 3 meters is acceptable, but when providing street level directions it is not.

For those use cases that do not require high accuracy, common ways of dealing with reducing the size of spatial data include degrading precision by reducing the number of decimals, and simplifying geometries using a simplification algorithm such as Ramer-Douglas-Peucker [37, 14] or Visvalingam–Whyatt [42]. These methods result in lower accuracy and precision.

Big spatial data is often not vector data, i.e., a representation of spatial things using points, lines, and polygons [20], but coverage data, i.e., gridded data: a data structure that maps points in space and time to property values [27]. For example, an aerial photograph can be thought of as a coverage that maps positions on the ground to colors. A river gauge maps points in time to flow values. A weather forecast maps points in space and time to values of temperature, wind speed, humidity and so forth. For coverage data, other methods are required to manage size.

The Data on the Web Best Practice recommends to provide 1) bulk download and 2) subsets of data. Providing bulk-download or streaming access to data is useful in any case and is relatively inexpensive to support as it relies on standard capabilities of web servers for datasets that may be published as downloadable files stored on a server. Subsets, i.e., extracts or "tiles", can be provided by having identifiers for conveniently sized subsets of large datasets that web applications can work with [10]. Effectively, breaking up a large coverage into pre-defined lumps that you can access via HTTP GET requests is a very simple API. A second way of supporting extracts, more appropriate for frequently changing datasets, is by supplying filtering options that return appropriately sized subsets of the specific dataset.

3.9. Crawlability

Search engines use crawlers to update their search index with new information that has been found on the Web. Traditional crawlers consist of two components: URL extractors, i.e., HTML crawlers that extract links from HTML pages in order to find additional sources to crawl, and indexers, i.e., different types of indexes, typically using the occurrence of text on a web page, that are maintained by search engines.

A major issue with crawlers identified in the early 2000's by Bergman [9] was the inaccessibility of the so-called Deep Web: information that was hidden to traditional crawlers as it is only accessible through services (e.g., forms) that require user input. Several solutions have since been introduced to access and index information on the Deep Web [17, 32].

Spatial Data services typically make information available only after user input has been provided, leading to a similar problem, i.e., a Deep Spatial Web. Further, these services are built to be accessed and searched by domain-specific applications rather than general web services and/or search engine crawlers. For example, in the OGC architecture, catalog services are intended to be used for searching spatial assets, not by general purpose search engines. However, a typical web user does not know that these catalogs exist and is accustomed to using general purpose search engines for finding information on the Web. Therefore, making sure that spatial data is indexable by web search engines is an important approach for making spatial data discoverable by users directly.

The addition of structured data to web services that are otherwise not accessible to search engines increases the visibility of a service or dataset in major search engines [16]. Schema.org, a single schema across a wide range of topics that includes people, places, events, products, offers, etc., and widely supported by Bing, Google, Yahoo! and Yandex, is the predominant way of marking up content and services on the Web with structured data to improve the presentation of the result in a search engine [15]. Google's Structured Data testing tool shows if schema.org markup on a web page is recognized. Experimental work such as Geonovum's Spatial Data on

the Web Testbed[7] describes ways to make spatial data indexable by publishing an HTML web page for the spatial dataset and each spatial thing that it contains; by using structured data, schema.org and links, as well as publishing XML sitemaps containing links to all data resources for spatial data services. Another example is the Dutch geoportal PDOK.nl which extensively publishes dataset metadata, for example the national roads dataset[8], resulting in better accessibility through common search engines. Several examples of spatial things published in this way are provided in Best Practice 2.

At present, spatial information, even when published in accordance with these guidelines, is not widely exploited by search engines. However, by increasing the volume of spatial information presented to search engines, and the consistency with which it is provided, it is expected that search engines to begin offering spatial search functions. Evidence is already seen of this in the form of contextual search, such as prioritization of search results from nearby entities. In addition, search engines are beginning to offer more structured, custom searches that return only results that include certain schema.org types, like Dataset, Place or City.

3.10. Other aspects

Spatial things can be shared with 0 to 3 or more dimensions. It may be difficult to combine similar things if the dimensions they are portrayed in differ. Although the best practice does not address this at length, it does recommend describing the number of dimensions in metadata and notes that one of the selection criteria for choosing a geometry format on the Web is the dimensionality of the data.

With some spatial data, the symbology associated with the data is of high importance because it communicates meaning. However, anything relating to rendering spatial data as maps (e.g., portrayal), such as symbology, was explicitly out of scope.

4. Gaps in current practice

The best practices described in brief in section 3, and in full in the Spatial Data on the Web Best Practice document, are compiled based on evidence of real-world application. This is in line with the fourth principle described in section 2 of this paper. How-

ever, there are several issues that inhibit the use or interoperability of spatial data on the Web, for which no evidence of real-world applied solutions was found. These issues are denoted "gaps in current practice" and described as such in this section. An issue is considered a gap when no evidence of real-world applied solutions in production environments was found. The term 'production environment' signifies a case where spatial data has been delivered on the Web with the intention of being used by end users and with a quality level expected from such data. In contrast, a "testing environment" is published with the intent of being tested so that bugs can be discovered and fixed and an experimental publication of spatial data on the Web is published with the intent of, for example, exploring possibilities, learning about the technology, or other goals besides publishing with the intent of serious use.

In the case of gaps, there might be emerging practice, i.e., a solution that has been theorized for a certain issue and has possibly been experimented on in testing / beta settings, but not in production environments. Gaps and emerging practices in the area of publishing spatial data on the Web are discussed in this section.

4.1. Representing geometry on the web

Location information can be an important 'hook' for finding information and for integrating different datasets. There are different ways of describing the location of spatial things: referencing the name of a well-known named place, describing a location in relation to another location, or providing the location's coordinates as a geometry. The latter allows the integration of data based on location using spatial reasoning, even when explicit links between things are not available, as well as, of course, showing spatial things on a map.

There are several aspects to representing geometries on the Web. First, there is the question of different serialization formats to choose from. In general, the formats are the same as for publishing any other data on the Web: XML, JSON, CSV, RDF, etc. How to select the most appropriate serialization is described in general terms in the Data on the Web Best Practice. As described in section 3.1, a single best way of publishing geometries was not identified in the Best Practice. With regards to geometries in particular, there are currently two reference geometry formats widely used in the geospatial and web communities respectively, GML [36] and GeoJSON [12].

---

[7] https://github.com/geo4web-testbed/general
[8] https://data.pdok.nl/datasets/nationaal-wegenbestand-wegen

GML provides the ability to express any type of geometry, in any coordinate reference system, and up to 3 dimensions (from points to volumes). GeoJSON supports only one coordinate reference system (WGS84), and geometries up to 2 dimensions (points, lines, surfaces). However, this is sufficient for the majority of use cases, and GeoJSON is more compact [30] as well as popular in the web community. In order to facilitate the use of geometry data on the Web, it is therefore desirable that GML-encoded geometries are made available also in GeoJSON, by applying not only the required coordinate reference system transformation, but, if needed, by simplifying the original geometry (e.g., by transforming a 3D geometry in a 2D one).

A second, related aspect is whether to publish geometries on the Web in self-contained files such as GML or GeoJSON, or rather to embed geometries as structured data markup in HTML, or in an RDF based way, i.e., as Linked Data. Choosing between these approaches - or not choosing but rather offering a combination of these - depends largely on the intended audience. The options for embedding geometry in HTML are limited. Typically this is done using schema.org as Microformats or JSON-LD, but used only for simple 0D-2D geometries (points, lines, surfaces), e.g., the centroid and/or 2D bounding box. As explained in section 4.1 dealing with crawlability, the advantage is that HTML with embedded data is indexed by search engine crawlers. RDF-based publication of geometry data is the most advanced option, but the audience for this is smaller than the others. GeoSPARQL offers a vocabulary that allows serialization of geometry as WKT or GML.

Third, there is a question of how to make geometries available in different CRSs. Section 3.2 explains the existence of many CRS and why spatial data should be published in CRSs that are most common to potential users. It follows that, on the one hand, the CRS should be specified for geometries published on the Web and, on the other hand, users should be able to find out which CRSs are available and how to get geometries using the CRS of their choice. It is common for a CRS to be described by its EPSG code. EPSG is a register of CRS maintained by the EPSG, a oil industry organization. Sometimes the CRS used is clear from the representation, as is the case with GeoJSON and schema.org, which only support one. In other cases the CRS needs to be specified either on the dataset level or the instance level. How this is done differs for each serialization. For example, in GeoSPARQL this is added as a prefix of the WKT literal while in GML an attribute *'srsName'* can be specified on geometry elements. In an OGC WFS request, users can specify the CRS they wish to use by specifying the *srsName* parameter. In GeoSPARQL the *getSRID* function returns the spatial reference system of a geometry, thus making it possible to request a specific CRS at a (Geo)SPARQL endpoint. However, these options require the user to be proficient in either Geospatial web services or Linked Data. A best practice for returning geometries in a specific requested CRS has not yet emerged. Many options can be found in current practice, including creating CRS-specific geometry properties (for example, the Dutch Land Registry does this), and supporting an option for requesting a specific CRS in a convenience API; but one best practice cannot yet be identified. Another option worth exploring might be the use of content negotiation, i.e., negotiate CRS as part of the content format for the geometry, as has also been proposed for encoding format. For example, this could be done with an extra media type parameter (e.g., application/ttl; geomLiteral="WKT"; crs="CRS84") or by adding specific request and response headers for negotiating CRS to the HTTP protocol.

However, providing different CRS might be too complicated to handle in the HTTP protocol. For example, multidimensional datasets will in general use multiple CRSs (e.g., horizontal, vertical and temporal, maybe more), and conversion between CRSs will in general introduce errors, so data in one CRS are not exactly the same as data in another CRS. Furthermore, CRS is just one of a number of parameters that may characterize a particular geometric representation of a Spatial Thing, including the type of geometry, its relationship to the Thing, method of interpolation, scale or resolution. However, offering a choice between all these parameters of data objects such as geometries might be an overloading of HTTP content negotiation protocols. It might therefore be more appropriate to handle this in the application layer.

4.2. A spatial data vocabulary

Although a large amount of geospatial data has been published on the Web, so far there are few authoritative datasets containing geometrical descriptions of their boundaries. Their number is growing (e.g., at the time of writing there are three authoritative spatial datasets publicly available as linked data in the Netherlands containing topographic, cadastral, and address data), but currently there is no common

practice in the sense of the same spatial vocabulary being used by most spatial data publishers. The consequence is the lack of a baseline during the mapping process for application developers trying to consume specific incoming data. Datasets describing administrative units, points of interest or postal addresses with their labels and geometries, and identifying these features with URIs could be beneficial not only for georeferencing other datasets, but also for interlinking datasets georeferenced by direct and indirect location information.

Direct georeferencing of data implies representing coordinates or geometries and associating them to a CRS. This requires vocabularies for geometries and CRSs. Further, indirect georeferencing of data implies associating them to other data on named places. Preferably, these data on named places should be also georeferenced by coordinates in order to serve as the basis for data linking between indirectly and directly georeferenced datasets. Vocabularies developed for representing specific sets of geospatial data on the Web should reuse as much as possible existing ones. This is the case for the vocabularies developed by IGN France for geometries[9], topographic entities[10], and CRS[11]. These vocabularies contain alignments with existing vocabularies, e.g., the class *geom:Geometry* is a subclass of both *sf:Geometry* (OGC simple features vocabulary) which is a subclass of the *Geometry* class of the GeoSPARQL vocabulary; and *ngeo:Geometry* (Neogeo vocabulary). Furthermore, the topographic entity class from the IGN France vocabulary is declared equivalent to the *Feature* class from the Geonames vocabulary.

In W3C basic Geo, it is assumed that the CRS used is WGS84. However, publishers might have data in a different, local CRS. Thus, there is need for a more generic class for, for example, a point geometry with the benefit of choosing the CRS of the underlying data [5]. Furthermore, there is need for a vocabulary to describe CRSs in a machine readable form; IGN France has created such a vocabulary.

Table 1 presents the implementation of the datasets containing complex geometries for the French territory [4], based on the implemented vocabularies at IGN France available at the endpoint http://data.ign.fr/id/sparql.

---

[9] http://data.ign.fr/def/geometrie/20160628.en.htm
[10] http://data.ign.fr/def/topo/20140416.en.htm
[11] http://data.ign.fr/def/ignf/20160628.en.htm

Table 2 Evidence of spatial data published by IGN-France.

Dataset	Named Graph	#NbTriples
Regions from IGN	http://data.ign.fr/id/geofla/region/	428305
CRS registry at IGN	http://data.ign.fr/id/ignf/	24062
Departments from IGN	http://data.ign.fr/id/geofla/departement/	405562
Arrondissements from IGN	http://data.ign.fr/id/geofla/arrondissement/	7245
Cantons Geofla	http://data.ign.fr/id/geofla/canton/	1658342
Communes from IGN	http://data.ign.fr/id/geofla/commune/	1061267

Vocabularies like the one by IGN France are created because, currently, no single standardized vocabulary is available that covers all needs. A possible way forward is an update for the GeoSPARQL spatial ontology. This would provide an agreed spatial ontology, i.e., a bridge or common ground between geographical and non-geographical spatial data and between W3C and OGC standards, conformant to the ISO 19107 [21] abstract model and based on existing available ontologies such as GeoSPARQL, the W3C basic geo vocabulary, NeoGeo or the ISA Core Location vocabulary.

This vocabulary would define basic semantics for the concept of a reference system for spatial coordinates, a basic datatype, or basic datatypes for geometry, how geometry and real world objects are related and how different versions of geometries for a single real world object can be distinguished. For example, it makes sense to publish different geometric representations of a spatial object that can be used for different purposes. The same object could be modelled as a point, a 2D polygon or a 3D polygon. The polygons could have different versions with different resolutions (generalization levels). And all those different geometries could be published with different coordinate reference systems.Thus, the vocabulary would provide a foundation for harmonization of the many different geometry encodings that exist today.

4.3. Spatial aspects of metadata

Even if all spatial data should become findable directly through search engines, data portals would still remain important hubs for data discovery - for exam-ple, because the metadata records registered there can be made crawlable. But, in addition, different data portals can harvest each others' information provided there is consistency in the types and meaning of included information, even if structures and technologies vary. Discovery of spatial data is improved in the Netherlands, for example, because the national general data portal harvests the spatial data portal and thus all spatial datasets are registered in the general data portal as well.

In the eGovernment sector, DCAT is a standard for dataset metadata publication and harvesting this metadata is implemented by eGovernment data portals. Because DCAT is lacking in possibilities for describing some specific characteristics of spatial datasets, an application profile for spatial data, GeoDCAT-AP, has been developed in the framework of the ISA Programme of the European Union, with the primary purpose of enabling the sharing of spatial metadata across domains and catalogue platforms.

To achieve this, GeoDCAT-AP defines RDF bindings covering the core profile of ISO 19115 [23] and the INSPIRE [19] metadata schema, enabling the harmonised RDF representation of existing spatial metadata. The focus was on the most used metadata elements, whereas additional mappings could be defined in future versions of the specification, based on users' and implementation feedback.

One of the outcomes of the development of GeoDCAT-AP was the identification of gaps in existing RDF vocabularies for representing some spatial information - such as coordinate reference systems and spatial resolution. But it also highlighted a key issue for spatial data, in that the use of global and persistent identifiers is far from being a common practice. Apart from making it difficult to implement a Linked

Data-based approach, this situation has negative effects in the geospatial infrastructure itself. E.g., it makes it impossible to unambiguously identify a spatial thing or a dataset over time, and it prevents an effective implementation of incremental metadata harvesting in a federated spatial data infrastructure (such as the INSPIRE one).

GeoDCAT-AP is referenced in the Spatial Data on the Web Best Practice as a possible approach for expressing spatial metadata on the Web, but is not yet an accepted standard. The next step is the evolution towards a single standard for metadata as it is used in data portals, without loss of relevant metadata, while still understandable and not too complicated. A working group in the Open Geospatial Consortium is currently working on this.

4.4. Describing dataset structure and service behaviours

Datasets may be arbitrarily large and complex, and may be exposed via services to expose useful resources, rather than a "download" scenario. Data gathered using automated sensors in particular may be impossible to download in its entirety due to its dynamic nature and potential volumes. It is therefore necessary in these cases to be able to adequately describe the structure of such data and how services interact to expose subsets of it - even individual records in a Linked Data context. Such datasets are common in the information processing world, and commonly organised in "hypercubes" - where "data dimensions" are used to locate values holding results.

A standard based on this dimensional model of data is the RDF Data Cube vocabulary [11]. It has been used to publish sensor data; for example to publish a homogenized daily temperature dataset for Australia over the last 100 years [29]. However, RDF Data Cube is lacking in possibilities for describing spatio-temporal aspects of data, which are very important for observations. One of the work items in the Spatial Data on the Web working group is therefore an extension to the existing RDF Data Cube ontology to support specification of key metadata required to interpret spatio-temporal data, called QB4ST.

QB4ST is an extension to RDF Data Cube to provide mechanisms for defining spatio-temporal aspects of dimension and measure descriptions. It is intended to enable the development of semantic descriptions of specific spatio-temporal data elements by appropriate communities of interest, rather than to enumerate a static list of such definitions. It provides a minimal ontology of spatio-temporal properties and defines abstract classes for data cube components (i.e., dimensions and measures) that use these, to allow classification and discovery of specialized component definitions using general terms.

QB4ST is designed to support the publication of consistently described re-usable and comparable definitions of spatial and temporal data elements by appropriate communities of practice. One obvious such case is the use of GPS coordinates described as decimal latitude and longitude measures. Another example is the intended publication of a register of Discrete Global Grid Systems (DGGS) by the OGC DGGS Working Group. QB4ST is intended to support publication of descriptions of such data using a common set of attributes that can be attached to a property description (extending the available RDF-QB mechanisms for attributes of observations).

4.5. Versioning of spatial data

Future Internet technologies will aim more and more to capture, make sense of and represent not only static but dynamic content and up-to-date information. Through internet technology, connections between devices and people will be realized through the exchange of large volumes of multimedia and data content. When the communication latency becomes lower, and the capacity of the communication gets higher with fifth generation (5G) mobile communications systems, the Internet will be an even more prominent platform to control real and virtual objects in different parts of our lives, such as healthcare, education, manufacturing, smart grids, and many more [28]. The Internet of Things [7] and Tactile Internet [38] are some of the technologies that aim to facilitate the interaction between people and devices, observe near real-time phenomena and actuate devices or robots. The Tactile Internet is focused on speeding up this interaction process and reducing the latency in communication systems.

Such high-speed communication will bring new challenges for intelligent systems. There is a big gap in the lightweight and semantically rich representation of versioning and temporal aspects of spatial data content. There have been a few attempts to represent changing and moving spatial objects, such as TimeSeriesML [12] and Moving Features [13]. However, although these ontologies provide a reasonably good semantic coverage, there is still a need for the devel-

---

[12] http://www.opengeospatial.org/standards/tsml
[13] http://www.opengeospatial.org/standards/movingfeatures

opment of lightweight and semantically rich representations to conduct enhanced (near) real-time operations. Having heavy semantic expressivity in ontologies can cause a burden on reasoning engines and can slow down the processing time for machines.

For instance, it is important to identify and represent the direction and coverage area of a surveillance camera or the orientations and positions of objects or people in the observed environment. In the future, a broken car will be fixed by a robot, or surgeries will be carried out by multiple robots using Tactile Internet [2]. It can also be envisioned that people, who have difficulty in walking, will not need to use a walking stick but merely a strap of an exoskeleton. These must be controlled by wireless systems to monitor the coverage, direction, and identify the objects and people around them including their shapes.

To conduct such activities, a better representation is needed for not only spatial information but also temporal and geometrical aspects of objects. Observed objects can change their size during the actuation process. For instance, a group of surgical robots will need to know about the shape of an organ and the changes regarding its size, geometrical shape, and orientation during surgery and exchange this information among themselves and with doctors to conduct an operation with high precision and low latency. A self-driving wheelchair or a self-driving car will be able to communicate with other sensor objects regarding the surrounding environment and direction to avoid obstacles. This will prevent possible accidents and harm caused by machines, such as not falling down stairs or running into objects with high speed or force.

In all these scenarios, lightweight representation and exchange of temporal-spatial knowledge are essential to understand and react fast enough to prevent disasters or to control the movement of devices. Having the means to represent the semantics of activities and phenomena at such high granularity and lightweight format will play a pivotal role in the development of future internet technologies. Moreover, it will allow machines to instantly exchange information including spatial and geometrical knowledge and carry out their tasks with high precision.

5. Conclusions

Spatial data has become ubiquitous with the explosive growth in positioning technologies attached to mobile vehicles, portable devices, and autonomous systems. It has proven to be fundamentally useful for countless things, ranging from everyday tasks like finding the best route to a location to solving the biggest global challenges like climate change adaptation. However, spatial data dissemination is heterogenous and although the Web is commonly used as a publication medium, the discovery, retrieval, and interpretation of spatial data on the Web is still problematic.

The Spatial Data on the Web Best Practices describe how linked data web principles can be applied to the world of complex spatial data to solve this problem. Good practices can be observed in current practice and have been collected into the Best Practices based on a set of principles and an examination of practice. In some cases a best practice has not yet emerged. There are still questions related to representing geometry on the Web, with regard to recommendable serialisation forms and formats, and the use of coordinate reference systems. A web-friendly way of publishing spatial metadata has not yet been described in full, especially with regards to the relevant subset of spatial metadata standards. A standardized ontology for spatial things that covers all needs for publishing spatial linked data is not yet available. Finally, there are new approaches emerging such as QB4ST, an extension to the RDF Data Cube to provide mechanisms for defining spatio-temporal aspects of dimension and measure descriptions. Notwithstanding these gaps and emerging solutions, a useful set of actionable best practices for publishing spatial data on the Web has been described. Following these guidelines will enable data users, web applications and services to discover, interpret and use spatial data in large and distributed web systems.

6. Acknowledgments

Bibliography

[1] C. Aggarwal, N. Ashish, and A. Sheth, The internet of things: A survey from the data-centric perspective. In Managing and mining sensor data (2013) pp. 383-428. Springer US.

[2] A. Aijaz, M. Simsek, M. Dohler, and G. Fettweis, "Shaping 5G for the Tactile Internet", 5G Mobile Communications, 2016.

[3] R. Albertoni and A. Isaac. Data on the Web Best Practices: Data Quality Vocabulary. W3C (2016). URL: https://www.w3.org/TR/2016/NOTE-vocab-dqv-20161215/

[4] G. A. Atemezing, Publishing and consuming geo-spatial and government data on the semantic web, PhD thesis (2015). At: http://www.eurecom.fr/fr/publication/4545/download/mm-publi-4545.pdf [accessed 2017-06-14]

[5] G. A. Atemezing and R. Troncy, Comparing Vocabularies for Representing Geographical Features and Their Geometry. In 5th International Terra Cognita Workshop, Boston, USA, 2012. 38, 40

[6] S. Athanasiou, L. Bezati, G. Giannopoulos, K. Patoumpas, and D. Skoutas. GeoKnow - Making the Web an Exploratory for Geospatial Knowledge. Market and Research Overview (2013). At http://svn.aksw.org/projects/GeoKnow/Public/D2.1.1_Market_and_Research_Overview.pdf [accessed 28 February 2017]

[7] P. Barnaghi and A. Sheth,"The Internet of Things: The Story So Far", IEEE Internet of Things, 2014.

[8] R. Battle and D. Kolas, Enabling the geospatial semantic web with Parliament and GeoSPARQ. Semantic Web, 3(4) (2012), p. 355-370.

[9] M. Bergman, The deep Web: Surfacing the hidden value. BrightPlanet (2012) At https://brightplanet.com/2012/06/the-deep-web-surfacing-hidden-value/ [accessed 20170614].

[10] D. Brizhinev, S. Toyer, and K. Taylor, "Publishing and Using Earth Observation Data with the RDF Data Cube and the Discrete Global Grid System", W3C Working Group Note 05 January 2017.

[11] R. Cyganiak, D. Reynolds, and J. Tennison. The RDF Data Cube Vocabulary. W3C (2012). http://www.w3.org/TR/vocab-data-cube/.

[12] H. Butler, M. Daly, A. Doyle, S. Gillies, S. Hagen, T. Schaub. RFC7946 The GeoJSON Format. (2016).

[13] R. Cyganiak, D. Wood and M. Lanthaler, RDF 1.1 Concepts and Abstract Syntax. W3C (2014). At: https://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/

[14] D. Douglas and T. Peucker, "Algorithms for the reduction of the number of points required to represent a digitized line or its caricature", The Canadian Cartographer 10(2), 112–122 (1973) doi:10.3138/FM57-6770-U75U-7727

[15] European Commission. GeoDCAT-AP: A geospatial extension for the DCAT application profile for data portals in Europe. 23 December 2015. At: https://joinup.ec.europa.eu/node/139283/

[16] R. V. Guha, D. Brickley, S. Macbeth. Schema.org: evolution of structured data on the web. CACM Volume 59 (February 2016) Issue 2, Pages 44-51.

[17] B. He, M. Patel, Z. Zhang, K. Chen-Chuan Chang. Accessing the deep web. Communications of the ACM. Volume 50 Issue 5 (2007), Pages 94-101.

[18] I. Jacobs and N. Walsh, Architecture of the World Wide Web, Volume One. W3C. 15 December 2004. W3C Recommendation. URL: https://www.w3.org/TR/webarch/

[19] INSPIRE Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community. [Online; accessed 2-February-2016] http://data.europa.eu/eli/dir/2007/2/oj

[20] ISO 19103:2015 Geographic information -- Conceptual schema language. ISO standard, ISO.

[21] ISO 19107:2003 Geographic information – Spatial schema. ISO standard, ISO.

[22] ISO 10109:2015 Geographic information – Rules for application schema. ISO standard, ISO.

[23] ISO 19110:2005 Geographic information – Methodology for feature cataloguing. ISO standard, ISO.

[24] ISO 19115-1:2014 Geographic information -- Metadata -- Part 1: Fundamentals. ISO standard, ISO.

[25] ISO 19117-1:2012 Geographic information -- Portrayal. ISO standard, ISO.

[26] ISO 19119-1:2016 Geographic information -- Services. ISO standard, ISO.

[27] ISO 19123:2005 Geographic information -- Schema for coverage geometry and functions. ISO/TC 211. ISO. 15 August 2005. URL: https://www.iso.org/standard/40121.html

[28] M. Jaber, M. A. Imran, R. Tafazolli, and A. Tukmanov, "5G Backhaul Challenges and Emerging Research Directions: A Survey." IEEE Access, 2016.

[29] L. Lefort, J. Bobruk, A. Haller, K. Taylor, and A. Woolf. A linked sensor data cube for a 100 year homogenised daily temperature dataset. In Proceedings of the 5th International Conference on Semantic Sensor Networks-Volume 904 (2012) pp. 1-16).

[30] W. Li, M. Song, B. Zhou, K. Cao, and S. Gao. Performance improvement techniques for geospatial web services in a cyberinfrastructure environment–A case study with a disaster management portal. Computers, Environment and Urban Systems, 54 (2015), p. 314-325.

[31] F. Maali, J. Erickson. Data Catalog Vocabulary (DCAT). W3C Recommendation. At: https://www.w3.org/TR/vocab-dcat/ [online; accessed 18 July 2017

[32] J. Madhavan, D. Ko, Ł. Kot, V. Ganapathy, A. Rasmussen, A. Halevy. Google's Deep Web crawl. Proceedings of the VLDB Endowment. Volume 1 Issue 2 (2008), Pages 1241-1252.

[33] I. Masser, All shapes and sizes: the first generation of National Spatial Data Infrastructure. International Journal Of Geographical Information Science Vol. 13 (1999), Iss. 1.

[34] J. Nogueras-Iso, F.J. Zarazaga-Soria, R. Béjar, P.J. Álvarez, and P.R. Muro-Medrano. "OGC Catalog Services: a key element for the development of Spatial Data Infrastructures." Computers & Geosciences 31.2 (2005): 199-209.

[35] M. Perry and J. Herring, OGC GeoSPARQL - A Geographic Query Language for RDF Data. At: http://www.opengeospatial.org/standards/geosparql (2010) [online; accessed 2 February 2017]

[36] C. Portele, Opengis geography markup language (gml) implementation specification v3.2.1, OGC document ref. no.: OGC (2007).

[37] U. Ramer, "An iterative procedure for the polygonal approximation of plane curves", Computer Graphics and Image Processing, 1(3), 244–256 (1972) doi:10.1016/S0146-664X(72)80017-0

[38] M. Simsek, A. Aijaz, M. Dohler, J. Sachs, and G. Fettweis, "5G-Enabled Tactile Internet", IEEE Journal on Selected Areas in Communications, 2016.

[39] K. Taylor and E. Parsons, Where is Everywhere: Bringing Location to the Web. IEEE Internet Computing 19(2):83-87 (2015)

[40] W. Tobler "A computer movie simulating urban growth in the Detroit region". Economic Geography, 46(2) (1970) p. 234-240.

[41] L. van den Brink, P. Janssen, W. Quak, and J. Stoter, Linking spatial data: semi-automated conversion of geo-information models and GML data to RDF. International

Journal of Spatial Data Infrastructures Research, 9 (2014), p. 59-85.

[42]   M. Visvalingam and J. Whyatt. Line generalisation by repeated elimination of points. The Cartographic Journal. Volume 30, issue 1, p. 46-51. 1993.