Publishing the Marine Regions Gazetteer as a Linked Data Event Stream

Britt Lonneville[1], Harm Delva[2], Marc Portier[1], Laurian Van Maldeghem[1], Lennert Schepers[1], Dias Bakeev[1], Bart Vanhoorne[1], Lennert Tyberghein[1] and Pieter Colpaert[2]

[1] *Flanders Marine Institute, Wandelaarkaai 7, 8400 Oostende, Belgium*
[2] *IDLab, Department of Electronics and Information Systems, Ghent University – imec, Technologiepark-Zwijnaarde 126, 9052 Gent*

Abstract

Marine Regions provides a standard of marine georeferenced locations, boundaries and regions for scientific and educational purposes as part of the LifeWatch project. To this end, Marine Regions creates, shares and maintains a hierarchical gazetteer. This gazetteer serves as a geographical backbone for a wide range of users, from biodiversity databases such as the World Register of Marine Species (WoRMS) over global fisheries initiatives such as Global Fishing Watch, to the maritime intelligence of MarineTraffic. While there is a large diversity in how Marine Regions data are used, we wish to minimize the number of APIs we provide in order to reduce the maintenance burden. In this paper, we demonstrate how we have solved this by introducing i) the Marine Regions ontology described using Linked Data, ii) a mapping of Marine Regions to Linked Data in subject pages; and iii) a Linked Data Event Stream (LDES) that can be used for replication and synchronization. These contributions allow us to focus on the open, semantic and meaningful publication of our data, and interested parties can then build APIs on top of the Event Stream, or derive useful subsets (e.g., based on geographical location). As part of this effort, Marine Regions entities are now described following the Linked Open Data principles and are available in various common formats through content negotiation. Additionally, the geometries of each data object are now directly accessible, without an extra web service call to an external OGC service.

Keywords

Semantic Web, Marine, Gazetteer, Linked Data, Open Science, Event Stream

1. Introduction

Digital gazetteers are well-placed instruments to bridge the gap between the informal way people refer to locations (often by name) and the formal world of communicating with and between machines [1]. A (digital) gazetteer, in its most elementary form, is a list of places [2].

S4BioDiv 2021: 3rd International Workshop on Semantics for Biodiversity, held at JOWO 2021: Episode VII The Bolzano Summer of Knowledge, September 11–18, 2021, Bolzano, Italy
britt.lonneville@vliz.be (B. Lonneville); harm.delva@ugent.be (H. Delva); marc.portier@vliz.be (M. Portier); laurian.van.maldeghem@vliz.be (L. Van Maldeghem); lennert.schepers@vliz.be (L. Schepers); dias.bakeev@vliz.be (D. Bakeev); bart.vanhoorne@vliz.be (B. Vanhoorne); lennert.tyberghein@vliz.be (L. Tyberghein); pieter.colpaert@ugent.be (P. Colpaert)
0000-0002-5839-9005 (B. Lonneville); 0000-0001-8272-0754 (H. Delva); 0000-0002-9648-6484 (M. Portier); 0000-0003-0663-5907 (L. Van Maldeghem); 0000-0002-6145-8248 (L. Schepers); 0000-0002-2308-1363 (D. Bakeev); 0000-0002-6642-4725 (B. Vanhoorne); 0000-0003-0577-1691 (L. Tyberghein); 0000-0001-6917-2167 (P. Colpaert)

Apart from place names and coordinates, it can contain a wide range of additional geographical knowledge such as hierarchies and translations [3]. Digital gazetteers are being used in many fields of research and for many applications, from historical data preservation [4] to natural language processing [5].

The Marine Regions system, part of the LifeWatch project[1], aims to provide a digital gazetteer primarily focused on the marine environment. Originally targeted on the river Scheldt, and the Belgian part and southern bight of the North Sea, it now contains over 60,000 features worldwide. Entities in the Marine Regions gazetteer are characterized by their unique identifier, the Marine Regions Geographic Identifier (MRGID). This identifier allows users to unambiguously refer to a Marine Regions entity and is persistent and resolvable. In this way, the gazetteer records can be easily integrated into larger knowledge networks, such as the World Register of Marine Species (WoRMS) [6], Global Fishing Watch [7], and MarineTraffic[2].

The Marine Regions gazetteer integrates data from multiple sources, including various authoritative marine gazetteers and ecological classifications (e.g., Marine Ecoregions of the World [8]). Currently, the top five external sources are the BGN Advisory Committee on Undersea Features[3] (ACUF), Canadian Geographical Names Data Base[4] (CGNDB), the GeoNames.org geographical database[5], the IHO-IOC GEBCO Gazetteer of Undersea Feature Names[6] and the SCAR Composite Gazetteer of Antarctica[7] (CGA).

Marine Regions entities can either have point, line or polygon geometries. This geometry is described by the derived centroid and (where available) bounding box coordinates. If the original source provides a more detailed geometric representation through its web services, this is visualized on the entity page. Furthermore, the gazetteer integrates additional information for its records (Figure 1). An entity's `PlaceType` describes the type of the object and can be administrative (e.g. `Exclusive Economic Zone`, `Marine Protected Area`) or physical (e.g. `seamount`, `ridge`). In addition, one or more `GeoNames` can be stored in every record, allowing us to deal with synonyms or names in different languages. The gazetteer also provides a hierarchy between the different entities, based on a parent-child relation (`partOf`). Other relations with Marine Regions entities can be described in the Marine Regions gazetteer, these can cover topological (`adjacentTo`, `streamsThrough`,...) and non-topological (`administrativePartOf`, `influencedBy`, ...) links. Other information that can be added to an entity includes relevant URLs and notes.

Data entry in the Marine Regions gazetteer follows a bottom-up approach [9]. Apart from the collection and integration of external sources, a community of selected editors can introduce new entities into the database. These records are added on an ad hoc basis and can be verified by other Marine Regions editors or the Marine Regions data management team.

---

[1] https://lifewatch.be/

[2] https://www.marinetraffic.com

[3] https://geonames.nga.mil/gns/html/acuf.html, features available via the National Geospatial-Intelligence Agency's GEOnet Names Server (https://geonames.nga.mil/gns/html/index.html)

[4] https://www.nrcan.gc.ca/earth-sciences/geography/geographical-names-board-canada/about-canadian-geographical-names-database/9180

[5] https://www.geonames.org/

[6] International Hydrographic Organization-Intergovernmental Oceanographic Commission General Bathymetric Chart of the Oceans, https://www.gebco.net/data_and_products/undersea_feature_names/

[7] Scientific Committee on Antarctic Research, https://data.aad.gov.au/aadc/gaz/scar/

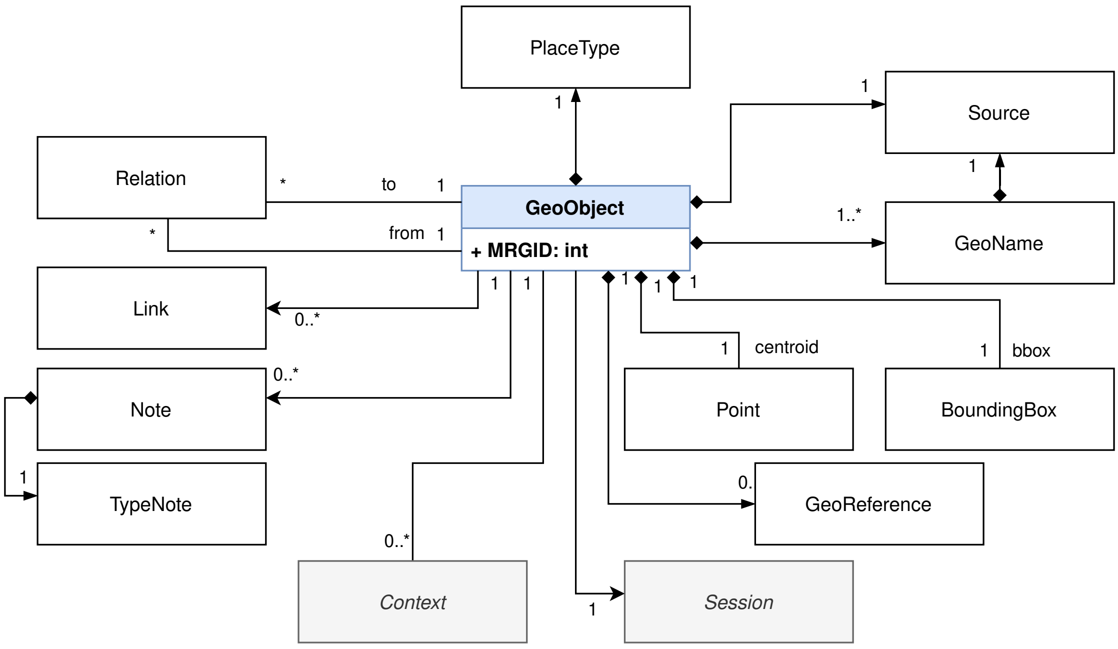

Figure 1: Data model of a Marine Regions entity.

Until recently, these data were was accessible only through the Marine Regions website, through OGC web services and through a JSON HTTP API. Starting from this situation, we wanted to solve two problems:

P1: Semantic Interoperability Despite the fact all the APIs are maintained by the same organization, they do not share a common semantic model and end up producing isolated views on top of the same data. While this caters for the many different expectations from existing clients, it does not further the more profound goal of interoperability. In this paper we report on how we aligned the Web APIs using a common Linked Data vocabulary.

P2: Maintainability of the Web ecosystem Maintaining a myriad of Web services on top of the Marine Regions gazetteer is labour intensive, which ultimately means less time can be spent on maintaining the data itself. In this paper we also report on how we implemented a Linked Data Event Stream as the base API for other APIs and clients.

This paper is structured as follows: section 2 gives an overview of how related (marine) gazetteers expose their data and the background information of Linked Data Fragments and Linked Data Event streams. Section 3 focuses on how we implemented these solutions for Marine Regions. Section 3.1 elaborates on how the semantic interoperability (P1) is accomplished. Section 3.2 focuses on the implementation of a Linked Data Event Stream as a solution for master data management on the Web (P2). Section 3.3 contains links to all relevant resources developed in this project. In the Discussion (Section 4) we highlight important design decisions that were made and their resulting limitations. And finally, section 5 summarizes lessons learnt and looks ahead to the future.

2. Related Work

To date, the way in which digital gazetteers and their data can be accessed varies wildly. For the five gazetteers that are used the most within Marine Regions, this ranges from plain data dumps (IHO-IOC GEBCO gazetteer, ACUF features, CGA, GeoNames.org), to custom-made APIs (IHO-IOC GEBCO gazetteer, ACUF features, CGNDB, GeoNames.org), OGC web services (ACUF features), to Linked Data endpoints (GeoNames.org). The GeoNames.org ontology was released in 2006 and has gradually improved since. Other global authoritative gazetteers, such as the Getty Thesaurus of Geographic Names[8] or the SeaVoX Salt and Fresh Water Body Gazetteer[9], are also adopting such a Linked Data approach in order to embed their data in the Semantic Web and facilitate Machine-to-Machine (M2M) communication. Providing digital gazetteers in such a way has the advantage of – among others – enhanced interoperability and facilitated data access [10]. Additionally, this also brings the data more into line with the FAIR Guiding Principles [11].

More specifically to the marine research community, ICES[10] provides a controlled vocabulary of codes, code types and lists that are used in their databases[11]. The vocabulary server of NERC[12] offers the widely-used BODC[13] terms [12] as Linked Data, where they utilize SKOS[14] [13] to define the semantics of various aspects of captured data: i) measurements, ii) units, iii) techniques and instruments, and iv) observable entities and their properties. Moreover, the SeaDataNet consortium publishes their metadata standards [14] as Linked Data through several SPARQL endpoints. Finally, the ambitious work of the ODIS[15] architecture group[16] bridges different local, regional and thematic nodes of marine and coastal data and information through Linked Open Data and Semantic Web technologies. By doing so, they are supporting full life-cycle management of (large scale) geospatial data [15] and proposing crosswalks connecting the oceangraphic knowledge graph [16].

Linked Data is often made accessible through data dumps, query APIs such as SPARQL, and de-referenceable subject pages. All these approaches can be laid out on the conceptual framework of Linked Data Fragments [17], as each exposes a fragment of the entire dataset. Data dumps and query APIs are the two extremes on the Linked Data Fragments axis[17]. This axis illustrates the trade-offs between different methods of publishing Linked Data on the Web. Data dumps put the data processing burden on the client's side but allow the most flexibility for clients. Query APIs put the processing burden on the server side but always restrict, in some

[8] https://www.getty.edu/research/tools/vocabularies/tgn/
[9] https://www.bodc.ac.uk/resources/vocabularies/seavox/
[10] International Council for the Exploration of the Sea, https://www.ices.dk
[11] https://vocab.ices.dk/
[12] Natural Environment Research Council, https://nerc.ukri.org/
[13] British Oceanographic Data Centre, https://www.bodc.ac.uk/
[14] Simple Knowledge Organization System, https://www.w3.org/TR/skos-reference/
[15] IOC Ocean Data and Information System (ODIS), http://odis.iode.org/
[16] https://github.com/iodepo/odis-arch
[17] Visualized at: https://linkeddatafragments.org

way, the way the data can be used. Beside this trade-off in the location of the data processing, there is also the trade-off on the up-to-dateness of the data. The optimal interval at which data dumps are produced needs to balance avoiding lagging behind on updates versus minimizing the total effort of producing the data dumps. Therefore, data dumps tend to be outdated soon after they are published. In contrast, users of a query API always have access to the system's current data.

Linked Data Event Streams (LDES) [18] are a specific kind of Linked Data Fragments, specifically for data consumers that wish to stay synchronized with a reference dataset. Rather than a fragmented collection of entities, an LDES is a collection of versions of entities – each representing an entity's state at a specific point in time. These versioned representations are gathered in a paginated append-only collection, marking full pages as immutable. This mechanism enables users to periodically check for new changes, which is more efficient than maintaining active connections to each data consumer [19]. Even third parties can thus stay synchronized with the base dataset and provide additional querying services that the first party did (or could) not provide themselves. This way, the LDES framework allows data publishers to prioritize which Web APIs to host on a certain budget, and which Web APIs other parties in the ecosystem can maintain.

3. Linked Open Data Publication of Marine Regions

3.1. Data Model

Related resources, such as those provided by NERC, make extensive use of the SKOS vocabulary – and a choice was made to do the same for the sake of interoperability. At its core, every `mr:GeoObject` is a `skos:Concept`, so that the Marine Regions dataset as a whole is a concept scheme. On top of this, we also chose to make as much use of the ISA Programme Location Core Vocabulary[18] as possible, as this focuses more specifically on addresses, geographic names and geometries and is consequently well-suited for gazetteer data. By defining all `GeoObjects` as both a `skos:Concept` and a `dct:Location`, we reconcile the conceptual content of the gazetteer entities with their geographical/geometric nature. The traditional SKOS predicates are used to name the entities, and a subproperty of `locn:geometry` is used to define the geospatial extents. This subproperty restricts the range to instances of `sf:Geometry`, defined by the OGC's SimpleFeatures vocabulary[19]. Although less evident from the used terms, we also found inspiration in the GeoNames.org ontology[20]. As in this ontology, we also make a distinction between a `mr:GeoObject`, and a `mr:Feature`. Whereas a `mr:GeoObject` is an entity that is contained in the Marine Regions dataset itself, all defined predicates refer to `mr:Feature` in their domain and range definitions. This is done so that reasoners do not assume that any entity that uses these predicates is part of Marine Regions itself. Figure 2 shows an overview of the used data classes, and how they relate to other vocabularies. Finally, to link to resources which do not have URIs, we make use of the `identifier` property from Schema.org[21]in combination with

[18]https://www.w3.org/ns/locn
[19]http://www.opengis.net/ont/sf
[20]https://www.geonames.org/ontology/documentation.html
[21]https://schema.org/

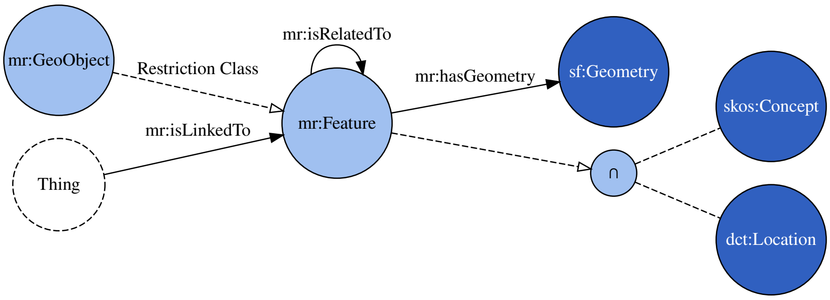

Figure 2: VOWL[23] visualization of the core of the used data model. A `mr:Feature` is defined to be both a `skos:Concept` and a `dct:Location`. An entity is only an instance of the `mr:GeoObject` class if it is explicitly defined as such.

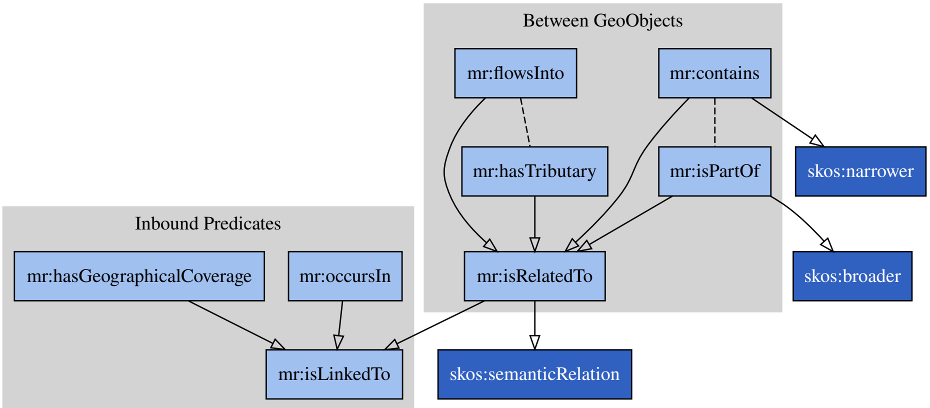

Figure 3: Schematic overview of the defined predicates to link from, or between, entities. The predicates to link between entities are subproperties of `skos:semanticRelation` and everything is a subproperty of `mr:isLinkedTo`

.

Wikidata identifiers[22]. For example, we use `http://www.wikidata.org/entity/P2326` to link to entities in the National Geospatial-Intelligence Agency's GEOnet Names Server.

The SKOS vocabulary contains predicates such as `skos:narrower` to model relations between different concepts. However, there are several ways one `GeoObject` may be more specific than another: it may be an administrative subdivision, it may be geospatially contained, and there may even be only some partial overlap between the two regions. To differentiate between the different ways two regions may be related, we defined nine subproperties of the existing SKOS predicates. These include hierarchical predicates and their inverses (e.g., `mr:isPartOf` and `mr:contains`), symmetric subproperties of `skos:related` (e.g., `mr:isAdjacentTo`), and subproperties of the general-purpose `skos:semanticRelation` (e.g., `mr:flowsThrough`).

To facilitate the reuse of the Marine Regions data, additional predicates are defined to refer

[22] https://www.wikidata.org/wiki/Wikidata:Identifiers
[23] Visual Notation for OWL Ontologies, http://vowl.visualdataweb.org

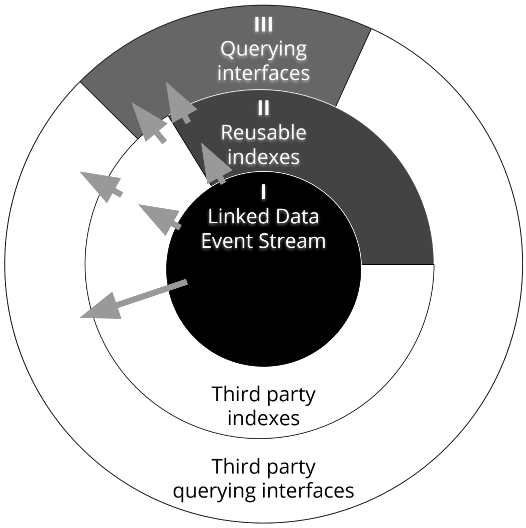

Figure 4: The layered design of an LDES architecture. The inner layer represents the raw data and is entirely the data publisher's responsibility. The other layers can be maintained by the publisher, or by third parties assuming they subscribe to the Event Stream themselves. There is no explicit expectation that third parties make their own services public, although Event Streams also facilitate ingesting the data into internal systems.

from any class of entity. Figure 3 graphically visualizes these predicates links and their relation to existing SKOS predicates. At the most general level, `mr:isLinkedTo` is defined as the predicate that may be used on any RDF *Thing*. Based on this predicate, three subproperties are defined:

- `mr:occursIn`: indicating a relation between a `GeoObject` and an entity that physically occurs within it, such as vessels, individuals of a species, the epicentre of an earthquake, the thermohaline circulation, or the oceanographic phenomenon of upwelling.
- `mr:hasGeographicalCoverage`: indicating a relation between a `GeoObject` and a non-physical entity, such as a dataset, data service or scientific publication.
- `mr:isRelatedTo`: indicating a relation between two `GeoObjects`, which acts as the superproperty for all predicates between `GeoObjects` (e.g., `mr:flowsThrough`).

3.2. The Linked Data Event Stream

As shown in figure 4, the idea of a Linked Data Event Stream (LDES) is to let data publishers focus on their core task, which is maintaining and publishing the data itself – while enabling third parties to replicate the entire dataset into whichever service they desire. In technical terms, an Event Stream is simply a collection of immutable objects. Each time the definition of a concept changes, this becomes a new object in the collection with its own URI. Because all objects in a certain page are guaranteed to be immutable, pages that are full can be labeled as immutable – enabling efficient caching of the entire collection. Conversely, data consumers can periodically poll the pages that are not labeled as immutable to discover changes to a dataset relatively quickly, depending on how often they poll the data.

Linked Data Event Streams show conceptual similarities with existing event streaming technologies, such as Apache Kafka[24] or even RSS newsfeeds[25]. We want to highlight what this publishing method can mean for scientific communities. As mentioned previously, users can ingest an Event Stream to set up services the data publisher did not develop. Examples of such services include (Geo)SPARQL endpoints, and document stores such as ElasticSearch to provide full-text search over the data. Moreover, since every change is published as a separate object, an Event Stream can be interpreted as a change log. Readers of an academic work that refers to a certain concept can thus infer the concept's state at the time of writing.

As shown in Listing 1, a distinction is made between the Event Stream and its published view(s). There are many ways an Event Stream may be paginated, which results in many possible views. And whereas the conceptual Event Stream contains the full change history, views can publish a subset by defining a *retention policy*. As the existing data infrastructure only stores the latest version of each entity, the published view's retention policy reflects this. The pagination strategy poses another challenge: the `dc:modified` timestamp is the most logical property to fragment on, but the changes are not spread out evenly in time. In fact, most gazetteer entities were updated very recently, even within a few days of each other, because we updated their stored geometries as part of this project. To overcome this challenge, each page only contains the version URIs, which clients can then dereference to access the version object itself. Keeping the page sizes small is crucial, as clients are expected to periodically poll these pages to discover updates.

3.3. Online Resources

The ontology is available at `http://marineregions.org/ns/ontology` and is maintained in a public Github repository[26]. There is also a code list of `PlaceTypes`, available at `http://marineregions.org/ns/placetypes`, which is generated from the database directly. The entities themselves are available at `http://marineregions.org/mrgid/{id}`, and listing 2 shows an extract of one of these entities. Finally, the LDES is available at `http://marineregions.org/feed`.

As is traditional, we make a distinction between the concepts (e.g., `http://marineregions.org/mrgid/3293`) and the documents describing the concept (e.g., `https://marineregions.org/mrgid/3293.ttl`) – with the concept URI redirecting to a relevant document, based on the request's `Accept` headers. Two RDF serialisations are currently supported, Turtle and JSON-LD, alongside the existing JSON, XML, and HTML options. The JSON-LD files are generated as regular JSON responses so that developers can choose to parse the data as RDF or as regular JSON.

[24]https://kafka.apache.org/
[25]https://rss.com/
[26]https://github.com/lifewatch/marineregions-ontology

```
<http://marineregions.org/feed?page=2021-05-10T15:00:00Z%2F2021-05-10T16:00:00Z>
  ldes:retentionPolicy [
    a ldes:LatestVersionSubset ;
    ldes:amount 1 ;
    ldes:versionKey ( dc:isVersionOf )
  ] .
<http://marineregions.org/feed>
  a ldes:EventStream ;
  tree:member <http://marineregions.org/mrgid/14?t=1620659836> ;
  tree:view
↪  <http://marineregions.org/feed?page=2021-05-10T15:00:00Z%2F2021-05-10T16:00:00Z>
↪  .
<http://marineregions.org/mrgid/14?t=1620659836>
  dc:isVersionOf <http://marineregions.org/mrgid/14> ;
  dc:modified '2021-05-10T15:17:16Z'^^xsd:dateTime .
```

Listing 1: Extract of one page in the LDES feed, which states that the `dc:isVersionOf` predicate is used to version the objects, and that only the most recent version is retained in this view of the Event Stream.

```
<http://marineregions.org/mrgid/3293>
  a mr:MRGeoObject, mrt:EEZ ;
  mr:hasGeometry
↪  <http://marineregions.org/mrgid/3293/geometries?source=4&attributeValue=3293>,
↪  <http://marineregions.org/mrgid/3293/geometries?source=79&attributeValue=3293> ;
  mr:isPartOf <http://marineregions.org/mrgid/14>,
↪  <http://marineregions.org/mrgid/2350> ;
  mr:isPreferredAlternativeOf <http://marineregions.org/mrgid/26567> ;
  mr:partlyContains <http://marineregions.org/mrgid/24676>,
↪  <http://marineregions.org/mrgid/28226> ;
  skos:prefLabel 'Belgian Exclusive Economic Zone'@en ;
  dcat:centroid '<http://www.opengis.net/def/crs/OGC/1.3/CRS84> Point(2.70504
↪  51.46545)'^^gsp:wktLiteral .
```

Listing 2: Extract of the Turtle representation of the Belgian Exclusive Economic Zone with relations and multiple geometries.

4. Discussion

Building a Linked Data strategy on top of an existing gazetteer that needs to maintain backward compatibility was a challenge. In this section we outline the design decisions we made, and the resulting limitations.

First of all, in order to have a single back end to serve the data from, an effort was made to bring the descriptions and the geospatial geometries, which are often sourced from third-party APIs, together in one database. However, some entities such as the Exclusive Economic Zones have highly detailed delineations, resulting in large amounts of data in the subject pages. In order to remediate this, the detailed geometries of a `GeoObject` are provided as additional links which the user can follow if needed. The subject pages do not only contain links to geometries, they also contain manually curated links to other related `GeoObjects`. However, some high-level entities (e.g., the Atlantic Ocean) are related to thousands of other entities. To avoid further inflating the subject pages with all these relations, each page contains at most 40 relations and the `hydra:next` predicate is used to point to subsequent pages, where the remaining relations can be found.

An additional complexity regarding the geometries is that some `GeoObjects` have multiple distinct geometries associated with them. This is because: (i) either the geometry of the object is a union of more than one feature (e.g., the Atlantic Ocean consists of the Atlantic Ocean *sensu stricto* plus its surrounding seas), or (ii) there are multiple possible demarcations and these are all displayed together (e.g., the Belgian part of the North Sea, where the demarcation depends on the landward baseline). For the Linked Data subject pages, the decision was made to list all these geometries separately. The user can then determine whether they prefer to consider the geometries separately, generate a spatial union, or select which geometry is most suitable for the application.

Publishing the Event Stream raised another issue: the existing gazetteer was designed to only contain the most recent version of every entity. Therefore, the Event Stream can only state when an entity was changed, and it does not disclose the actual cause of the change, nor does it provide an archive of all previous versions. However, it is compatible with existing LDES software libraries[27], so that additional services such as archiving can be created.

The Marine Regions data model (as shown in Figure 1) contains several elements which are currently not being disclosed through the Linked Data subject pages:

- The session element, which holds metadata of the user-session that modified the Marine Regions record in the gazetteer.
- The context of a `GeoObject`, which is a generic label that can be assigned to cluster entities into groups for various purposes (e.g., data management). Since this is mainly used for internal processes and the added value for external parties is uncertain, we decided to omit this information from the Linked Data publication.

Finally, our primary concern regarding the Marine Regions Geographic Identifiers is their persistence, as they have already been in use for years. This is reflected in the use of the HTTP

[27] Such as https://github.com/treecg/event-stream-client and https://github.com/Informatievlaanderen/ldes2service

scheme for the concept URIs, but this is also why a handle-based approach (as chosen for e.g., FAIR Digital Objects[28]) is currently not being considered.

5. Conclusion and Future Work

To raise the semantic interoperability (P1), we introduced the Marine Regions ontology and we mapped the existing subject pages to also expose the entity information as Linked Data. Alongside this, we improved the maintainability of our Web APIs (P2) by publishing the data as a Linked Data Event Stream from which other APIs can be created.

The Marine Regions gazetteer can now be accessed as Linked Data by dereferencing the subject pages with the appropriate `Accept` HTTP headers, or through our REST APIs [29]. Our own plans include using these APIs to improve our existing R library[30] and significantly extend its functionality. The Event Stream can be used[31] to replicate the full dataset in any back end. In future work, we will host our own LDES fragmentations so that users can be more selective in the data they choose to replicate. For example, a geospatial fragmentation should facilitate the replication of data from a specific country or continent.

We end with this quote we found inspirational for this work: "Semantics is a cornerstone in state-of-the-art data management, notwithstanding the specific domain; without semantics, we would helplessly drown in a deluge of unintelligible Big Data" [20].

References

[1] M. F. Goodchild, L. L. Hill, Introduction to digital gazetteer research, International Journal of Geographical Information Science 22 (2008) 1039–1044.

[2] P. Bol, M. F. Goodchild, H. Kerfoot, R. Shaw, R. Singh, K. Grossner, K. Janowicz, C. Kessler, R. Simon, L. Isaksen, et al., Placing names: Enriching and integrating gazetteers, Indiana University Press, 2016.

[3] D. Ahlers, Assessment of the accuracy of geonames gazetteer data, in: Proceedings of the 7th workshop on geographic information retrieval, 2013, pp. 74–81.

[4] P. Aucott, H. Southall, Locating past places in britain: Creating and evaluating the gb1900 gazetteer, International Journal of Humanities and Arts Computing 13 (2019) 69–94.

[5] M. Kejriwal, P. Szekely, Neural embeddings for populated geonames locations, in: International Semantic Web Conference, Springer, 2017, pp. 139–146.

[6] L. Vandepitte, B. Vanhoorne, W. Decock, S. Dekeyzer, A. Trias Verbeeck, L. Bovit, F. Hernandez, J. Mees, How aphia—the platform behind several online and taxonomically oriented databases—can serve both the taxonomic community and the field of biodiversity informatics, Journal of Marine Science and Engineering 3 (2015) 1448–1473.

[28] https://fairdo.org/

[29] https://www.marineregions.org/gazetteer.php?p=webservices&type=rest/

[30] https://github.com/ropensci/mregions

[31] Using tools such as https://github.com/TREEcg/event-stream-client/tree/main/packages/actor-init-ldes-client

[7] W. Merten, A. Reyer, J. Savitz, J. Amos, P. Woods, B. Sullivan, Global fishing watch: Bringing transparency to global commercial fisheries, arXiv preprint arXiv:1609.08756 (2016).
[8] M. D. Spalding, H. E. Fox, G. R. Allen, N. Davidson, Z. A. Ferdaña, M. Finlayson, B. S. Halpern, M. A. Jorge, A. Lombana, S. A. Lourie, et al., Marine ecoregions of the world: a bioregionalization of coastal and shelf areas, BioScience 57 (2007) 573–583.
[9] E. Acheson, S. De Sabbata, R. S. Purves, A quantitative analysis of global gazetteers: Patterns of coverage for common feature types, Computers, Environment and Urban Systems 64 (2017) 309–320.
[10] B. Regalia, K. Janowicz, G. Mai, D. Varanka, E. L. Usery, Gnis-ld: Serving and visualizing the geographic names information system gazetteer as linked data, in: European Semantic Web Conference, Springer, 2018, pp. 528–540.
[11] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, et al., The fair guiding principles for scientific data management and stewardship, Scientific data 3 (2016) 1–9.
[12] A. Leadbetter, R. Lowry, D. Clements, The nerc vocabulary server: version 2.0, in: Geophysical Research Abstracts, volume 14, 2012.
[13] A. Miles, B. Matthews, M. Wilson, D. Brickley, Skos core: simple knowledge organisation for the web, in: International conference on dublin core and metadata applications, 2005, pp. 3–10.
[14] A. Leadbetter, R. Thomas, Exposing the seadatanet metadata catalogues via sparql endpoints, Bollettino di Geofisica (2018) 37.
[15] M. Koubarakis, K. Bereta, G. Papadakis, D. Savva, G. Stamoulis, Big, linked geospatial data and its applications in earth observation, IEEE Internet Computing 21 (2017) 87–91. doi:10.1109/MIC.2017.2911438.
[16] M. Zarate, P. Rosales, G. Braun, M. Lewis, P. Fillottrani, C. Delrieux, OceanGraph: Some Initial Steps Toward a Oceanographic Knowledge Graph, 2019, pp. 33–40. doi:10.1007/978-3-030-21395-4_3.
[17] R. Verborgh, O. Hartig, B. De Meester, G. Haesendonck, L. De Vocht, M. Vander Sande, R. Cyganiak, P. Colpaert, E. Mannens, R. Van de Walle, Querying datasets on the web with high availability, in: International Semantic Web Conference, Springer, 2014, pp. 180–196.
[18] D. Van Lancker, P. Colpaert, H. Delva, B. Van de Vyvere, J. Rojas Meléndez, R. Dedecker, P. Michiels, R. Buyle, A. De Craene, R. Verborgh, Publishing base registries as linked data event streams, in: Proceedings of the 21th International Conference on Web Engineering, 2021.
[19] B. Van de Vyvere, P. Colpaert, R. Verborgh, Comparing a polling and push-based approach for live open data interfaces, in: International Conference on Web Engineering, Springer, 2020, pp. 87–101.
[20] G. Bordogna, C. Fugazza, P. Tagliolato Acquaviva d'Aragona, P. Carrara, Implicit, formal, and powerful semantics in geoinformation, ISPRS International Journal of Geo-Information 10 (2021) 330.