Semantic Web 0 (0) 1
IOS Press

Observational/Hydrographic data of the South Atlantic Ocean published as LOD

Marcos Zárate [a,b,*], Germán Braun [c,b] Mirtha Lewis [a] and Pablo Fillottrani [d,e]

[a] *Centre for the Study of Marine Systems, Centro Nacional Patagónico (CENPAT-CONICET), Argentina*
[b] *National Scientific and Technical Research Council (CONICET), Argentina*
[c] *Grupo de Investigación en Lenguajes e Inteligencia Artificial, Departamento de Teoría de la Computación, Universidad Nacional del Comahue (UNCOMA), Neuquén, Argentina*
[d] *Computer Science and Engineering Department, Universidad Nacional del Sur (UNS), Bahia Blanca, Argentina*
[e] *Comisión de Investigaciones Científicas, provincia de Buenos Aires, Argentina*

Abstract. This article describes the publication of occurrences of Southern Elephant Seals *Mirounga leonina (Linnaeus, 1758)* as Linked Open Data in two environments (marine and coastal). The data constitutes hydrographic measurements of instrumented animals and observation data collected during census between 1990 and 2017. The data scheme is based on the previously developed ontology *BiGe-Onto* and the new version of the *Semantic Sensor Network ontology (SSN)*. We introduce the network of ontologies used to organize the data and the transformation process to publish the dataset. In the use case, we develop an application to access and analyze the dataset. The linked open dataset and the related visualization tool turned data into a resource that can be located by the international community and thus increase the commitment to its sustainability. The data, coming from Peninsula Valdés (UNESCO World Heritage), is available for interdisciplinary studies of management and conservation of marine and coastal protected areas which demand reliable and updated data.

Keywords: Hidrographic & Observational Data, Linked Open Data, Semantic Sensor Network, BiGe-Onto

1. Introduction

In the ecology domain, research teams collect and store biological and environmental information over the years/decades in database systems to answer their own query. However, this information is isolated from other datasets for interoperating with and, in addition, is not ready to be accessible for machines. Particularly, in marine science the data collection is a process of cumulative logistic complexity, which makes it important to work on the curation and sustainability of the database, in both the short and long term. It is of great benefit for scientific institutions to publish their datasets following the *Linked Data principles* [1] not only for interlinking and easy cross-referencing but also for other purposes that are not foreseen at the moment of publication. The state of the art in the last decade shows that together with technology to collect data, the semantic interoperability has further grown in importance [2]. The dataset must be described by rich metadata in a findable resource, through controlled vocabularies and must be assigned a unique identifier.

This paper integrates observational and hydrographic datasets based on the SOSA/SSN ontology [3] and *BiGe-Onto* ontology [4]. As far as we know, this work is the first to publish linked open data occurrences of a specie in two geographical environments (coast and marine) collected along two decades. Data comes from a research program focused in SES in Patagonia Argentina "*Temporal and spatial distribution of the southern elephant seal colony in Península Valdés, Argentina*" [5]. The program started in 1990 to study demography and life history strategies of Southern Elephant Seals (SES) *Mirounga leonina (Linnaeus, 1758)*, together with the research of foraging areas and dive behavior, and contribute to understanding the effects on the species behind the changes in the ecosystem

*Corresponding author. E-mail: zarate@cenpat-conicet.gob.ar.

of SW Atlantic Ocean. The research site is located in Península Valdés (PV), which has been a UNESCO World Natural Heritage since 1999 [6].

During the annual cycle the SES come ashore to breed and molt. The rest of the year they are at sea, traveling long distances throughout its extensive migration (up to 8 months and 12000 km. of round trip), and diving continuously to a depth of 2000 meters or more. During their terrestrial phase they are also characterized by high fidelity to the site where they have previously been [7, 8]. The behavior during the marine phase shows that SES are ideal carriers of devices, providing physical profiles, (i.e. hydrographic of the water column). For tracking the SES at sea, researchers make use of miniaturized animal-attached tags for relaying data, known as biologging domain [9], and cover animal migration and oceanographic measurements [10]. The instruments deployed on the seal return, at a low cost, large volumes of hydrographic data in regions never studied directly by buoys or oceanographic vessels, and collecting large amounts of information associated with the key habitats in the South Atlantic Ocean.

This paper is organized as follows: Section 2 describes the SES database. Section 3, briefly presents the network of ontologies. Section 4, shows examples of how the data are organized. Section 5, describes the populating processes and the links to other datasets. Section 6, shows the application developed to access and analyze the dataset. Finally we conclude by presenting an analysis of our work and perspectives.

2. SES Database

Data are recorded from measurements of physical variables and locations obtained in two different stages. First stage involves an annual census, which takes place during the breeding season of SES. The second stage starts at the end of the breeding season, when SES go back at sea for foraging purposes. Below we briefly describe how data are generated and recorded in each stage. During the breeding season, the SES haul out to the beach to breed. Annual censuses on foot along the coast of the colony is an arduous but indispensable work to know distribution and trend of the population. The objective is to count each of the harems scattered on the beaches of the PV to determine the number of offspring born in a season. Counts carried out during 2-3 days at peak of the breeding season (October 3-7), when most of the population is ashore. All the breeding groups were counted and located along 200 km of coastline, divided into sections and each census taker is assigned to a route. The census taker must count for the number of animals and classify them by sex and age males, females and pups. Hereinafter, we will call the procedure of counting individuals in a certain place *Occurrence*. Each occurrence was georeferenced (latitude and longitude) and demographic data included date and time, group size and substrate where the SES is located. All information about these censuses is recorded in a field book and then uploaded into a MySQL database. Table 1 summarizes the most relevant fields for the conducted censuses.

Table 1
Main fields registered during a census.

Field	Description
Observer	Person in charge
Date	Date of observations
Start time	Census' starting hour
End time	Census' finalizing hour
Location	Latitude and longitude of the occurrence
Weather	Cloudy, windy, rainy, clear
Tide	High/low
Category	Seals' stage (adult, yeraling, pups, etc.)
Sex	Female, male, not determined
Count	Amount of SES observed

At the end of breeding, SES go back at sea for foraging. The trip is monitored by small computers designed and built by Wildlife Computers Inc.[1] with sensors to take measurements about their location and immediate environment. The instrument is deploy when the seal is on land before the migration into the sea begins. Time Depth Recorder (TDR) records, time, depth, and temperature every 30 minutes during round foraging trip. The position is also registered when the seal ascents to the surface. Table 2 summarizes the fields that are most relevant in diving.

The censuses and the deploy of the instruments are carried out by the research team belongs to *Centre for the Study of Marine Systems* hosted in Puerto Madryn, Patagonia Argentina (CESIMAR-CENPAT-CONICET)[2]. The institute is engaged in oceanographic and marine research activities, monitoring information management and data acquisition activities on different platforms.

[1] https://wildlifecomputers.com/
[2] https://cenpat.conicet.gov.ar/cesimar/

Table 2
Main fields registered during SES diving

Field	Description
SES identifier	4-character alphanumeric identifier
Dive number	Total dives in one journey
Start time	Dive starting time and date
End time	Dive finalizing time and date
Dive depth	Maximum, average and minimum (in meters)
Temperature	Bottom and surface (in Celsius degrees)
Location	Latitude and longitude recorded

3. Ontologies used to model observations and hydrographic profiles

The core of our ontologies network is composed by SOSA/SSN [3] and *BiGe-Onto* [4], which can be jointly used for both hydrographic profiles and observational data publication. These ontologies are linked to other ones describing different sub-domains, and thus creating such network. As a consequence, the resulting network is composed mainly by the following:

- an ontology to describe the sensors used to measure hydrographic profiles
- an ontology to describe the occurrences made during censuses
- an ontology to describe the associated measures.
- an ontology to describe the locations and places of interest
- an ontology to describe temporality of events
- an ontology to describe scientific publications

In this section, we briefly summarize these ontologies used for the publication of our dataset, indicating the reuse of concepts.

Semantic Sensor Network (SSN) Ontology: The Semantic Sensor Network (SSN) is a generic ontology related to sensor observations [11]. This ontology has been updated to become a W3C recommendation, and currently it is a lightweight one dedicated to sensor and actuator descriptions. It has been called Sensor, Observation, Sample, and Actuator (SOSA) pattern. The link between SSN and SOSA is described as follows in [3]. The classes we have reused from SOSA/SSN ontology are:

- *sosa:Observation*: to describe the measurements context.
- *sosa:FeatureOfInterest* to specify the observed phenomena. In our case, temperature and depth.
- *sosa:ObservableProperty* to specify the measured property of the observed phenomena (e.g. average depth).
- *sosa:Platform* to represent the platform hosting a sensor. In our cases, the platform is always the SES.
- *sosa:Sensor* to describe sensors hosted by platform (e.g. TDR).
- *sosa:Result* to represent the measurement values from the sensors.

We have also reused the main properties associated with these classes: *sosa:observedProperty*, *sosa:hosts*, *sosa:madeBySensor*, *sosa:hasFeatureOfInterest*, and *sosa:hasResult*.

BiGe-Onto Ontology: is an ontology designed for modeling Biodiversity and Marine Biogeography data [4]. The main concept of *BiGe-Onto* is an *occurrence*. Given that the censuses are occurrences of SES at a specific time and place, we consider *BiGe-Onto* fits the nature of our data. At the same time, it reuses different vocabularies such as Darwin Core (DwC) [12], which is the core one in *BiGe-Onto*. Its main classes are: *dwc:Occurrence*, *dwc:Event*, *dwc:Taxon* and *dwc:Organism*. Moreover, *BiGe-Onto* reuses *foaf:Person void:Dataset* and *dc:Location*, among others.

Since *BiGe-Onto* mainly describes occurrences, which dependent on other concepts to exist, we also outline below some of the most important properties defined for relating such occurrences:

- *bigeonto:associated*: each occurrence is described based on the existence of an organism at a particular place and at a particular time. Organisms are related to a taxon by means of *bigeonto:belongsTo*.
- *bigeonto:has_event*: occurrences take place during a sampling event at a location given by *bigeonto:has_location*, which is also characterized *bigeonto:caracterizes* by a specific environment. The relations between *bigeonto:Environment* and EnvO classes are primarily controlled by the Relations Ontology (RO)[3] respectively.
- *dwciri:recordedBy*: this property provides information about people, groups, or organizations, who have recorded the occurrence. It is also reused from the DwC URI namespace and en-

[3] https://github.com/oborel/obo-relations

ables non-literal ranges for its analogous with DwC, *dwc:recordedBy*.
– *dwciri:inDataset*: This object property is provided to link a subject dataset record to the dataset which contains it.

The Quantity, Unit, Dimension and Type (QUDT): is a collection of OWL ontologies and vocabularies [13]. The QUDT schema defines the base classes, properties, and restrictions used for modeling physical quantities, units of measure, and their dimensions in various measurement systems. QUDT also contains a set of vocabularies to define units for different domains. We have reused the unit vocabulary that categorizes units in different classes. This vocabulary also provides individuals to identify units such as *qudt:Meter* or *qudt:DegreeCelsius*.

GeoSPARQL Ontology: GeoSPARQL [14] is an Open Geospatial Consortium (OGC) standard for supporting the representation and querying of geospatial data on the Semantic Web. As such, it is based on the OGC's Simple Features model, with some adaptations for RDF. GeoSPARQL designates a vocabulary for representing geospatial data in RDF, and it defines an extension to the SPARQL query language[4] for processing them, together with both a small ontology for representing features[5] and geometries[6], and a number of SPARQL query predicates and functions. All these definitions are derived from other OGC standards so that they are well grounded and documented. Using the new standard should ensure two things: (1) if a data provider uses the spatial ontology in combination with an ontology of their domain, these data can be properly indexed and queried in spatial RDF stores; and (2) RDF-compliant triple stores should be able to properly process the majority of spatial RDF data. This ontology is used to describe the location of each occurrence, and the beaches involved. We reuse the classes *geo:Feature* and *geo:Geometry*, and the associated properties *geo:hasGeometry*, *geo:asWKT*, etc.

The W3C Time Ontology: The W3C Time ontology [15] enables the description of time instants and intervals. Hence it may be useful when we need to describe the timestamps or the time period associated with the measurements made by the observers o to the SES. We reuse the classes *time:Interval* and *time:Instant*, and the associated properties *time:hasEnd*, *time:hasBeginning*, *time:inXSDDateTimeStamp*, etc.

[4]https://www.w3.org/TR/rdf-sparql-query/

[5]A feature is simply any entity in the real world with some spatial location.

[6]A geometry is any geometric shape, such as a point, polygon, or line, and is used as a representation of a feature's spatial location.

FRBR-aligned Bibliographic Ontology (FaBiO): is an ontology [16] for recording and publishing on the Semantic Web descriptions of entities that are published or potentially publishable, and that contain or are referred to by bibliographic references, or entities used to define such bibliographic references. FaBiO classes are structured according to the FRBR schema of Works, Expressions, Manifestations and Items. Additional properties have been added to extends the FRBR data model by linking Works and Manifestations. We reuse the classes *fabio:Expression*, *fabio:JournalArticle*, *fabio:Article*, *fabio:Presentation*, *fabio:Book*, *fabio:BookChapter*, *fabio:Notebook*, *fabio:Dataset* and *fabio:ReportDocument* to identify the kind of document published. We also reuse the associated properties such as *prism:doi*, *prism:publicationDate*, etc.

Additionally, we have used vocabularies of the oceanographic domain as The Natural Environment Research Council (**NERC**) Vocabulary Server [17] supported by the British Oceanographic Data Center (BODC)[7], provides access to lists of standardized terms that cover a broad spectrum of disciplines of relevance to the oceanographic and wider community. A list of prefixes and their corresponding URIs are listed in Table 3.

4. Data model and URIs

Based on the network of ontologies described in the previous section, we are now able to create a dataset containing all the individuals describing hydrographic profiles and occurrences taken during the censuses. Now we explain the decisions taken in order to create resource URIs and we provide examples of resource descriptions.

4.1. Resource URIs for hydrographic profiles

This section presents the main URI design decisions and conventions used. Table 4 provides a summary of the main types of URIs that we generate. The first column presents the type of resources. The

[7]https://www.bodc.ac.uk/

Table 3
Reused vocabularies and ontologies.

Ontology/Vocabulary name	Prefix	URI
BiGe-Onto ontology	bigeonto	http://www.w3id.org/cenpat-gilia/bigeonto/
Semantic Sensor Network Ontology	ssn	http://www.w3.org/ns/ssn/
Sensor, Observation, Sample, and Actuator Ontology	sosa	http://www.w3.org/ns/sosa/
Darwin Core (literal values)	dwc	http://rs.tdwg.org/dwc/terms/
Darwin Core (IRI values)	dwciri	http://rs.tdwg.org/dwc/iri/
GeoSPARQL ontology	geosparql	http://www.opengis.net/ont/geosparql#
W3C Time Ontology	time	http://www.w3.org/2006/time#
FRBR-aligned Bibliographic Ontology	fabio	http://purl.org/spar/fabio#
NERC vocabulary server	nerc	http://vocab.nerc.ac.uk/collection/P06/

second column indicates the associated class which types the resources. The last column contains the name pattern used to generate the resource URIs. The base URI for our dataset is *http://linkeddata.cenpat-conicet.gob.ar/resource/*. Its prefix is *base*. Our generic name pattern to produce URIs for each object is Base URI +"/"+ nameOfClass +"/"+ objectIdentifier. The "_" character is used between two object identifiers. For example, the URI that represents the SES platform *AAEU* is: *base:platform/SES_AAEU*

4.1.1. Platform and Sensor

We have made the decision to consider SES as an oceanographic sampling platform. The individual that represents the platform is an instance of the *sosa:Platform* class. Each sensor hosted by the SES is represented by an instance of the class *sosa:Sensor*. Figure 1 presents the description of the TDR sensor. The TDR is identified by an URI generated using the sensor model, e.g *TDRMK3*. This URI is typed by the class *sosa:Sensor*. The *sosa:host* property links the *sosa:Platform* instance to the TDR URI. The sosa:observes property links the TDR URI to an instance of *sosa:ObservableProperty* that is labelled by the string *average depth*.

4.1.2. Observation

An observation describes the context of a measurement made by a sensor, in the case of TDRs, the measurements are: time, depth and temperature. Figure 2 represents an observation made by the TDR on the average depth at a given point. The properties *sosa:hasFeatureOfInterest*, *sosa:madeBySensor*, *sosa:observedProperty* and *sosa:hasResult* link our specific observation with its corresponding observed property, location, sensor and measurement value. We create an instance of *sosa:FeatureOfInterest* class that represents the depth phenomenon. GeoSPARQL is used to describe the precise location of the SES during the trip. As shown in Figure 2, the geometry of the trip made by the SES is a set of points expressed by a WKT string. This string is linked to a *geo:Geometry* instance by the *geo:asWKT* property. The *geo:hasGeometry* property links the *sosa:Platform* instance to an instance of the *geo:Geometry* class.

4.1.3. Phenomenon Time

TDR measures the depth in an instantaneous manner. Figure 3 presents an observation produced by the TDRMK3. The *sosa:phenomenonTime* property points to an instance of the class *time:Instant*. The property *time:inXSDDateTimeStamp* connects the *time:Instant* instance to an *xsd:dateTime* value. Sometimes a measurement is related to a period of time. For example, the TDR measures the duration of a dive during a immersion. The property *sosa:phenomenonTime* links the *sosa:Observation* to an instance of the class *time:Interval*. The properties *time:hasBeginning*, *time:hasEnd* and *time:hasDuration* specify respectively the beginning, the end and the duration of the interval.

4.1.4. Occurrence

To represent the SES visual observations made during a census, we use the class *dwc:Occurrence*. In Figure 4 you can see the observation of 25 juvenile females on August 22, 1990. The property *bigeonto:has_event* connects the instance of occurrence with the instance of the event *bigeonto:BioEvent*. In the same way, each event instance is related to an instance of the *geo:Geometry* class through the *bigeonto:has_location* relationship. On the other hand, the *dwciri:recordedBy* property relates the instance of the occurrence to instances of the *foaf:person* class that perform the observation. Finally, the occurrence is associated with an instance of the class *dwc:Organism* through the *bigeonto:associated* property and the or-

Table 4

URI generation templates for resources. First part of the table describes the patterns of URIs related to hydrographic profiles, while the second describes those related to observational data.

Object	Class	pattern
SES Platform	*sosa:platform*	*base:platform/SES_{sesID}*
TDR	*sosa:sensor*	*base:sensor/{TDR type}*
Feature of Interest	*sosa:FeatureOfInterest*	*base:featureOfInterest/NaturalPhenomenonLabel*
Observable Property	*sosa:ObservableProperty*	*base:observableProperty/NaturalPhenomenonLabel*
Instant	*time:Instant*	*base:instant/{date}T{time}*
Duration	*time:Duration*	*base:duration/{dive duration}*
Interval	*time:Interval*	*base:interval/{sesID}_{beginDive}_{endDive}*
Observation	*sosa:Observation*	*base:observation/{diveNo.}_of_{sesID}_on_{NaturalPhenomenon}*
Result value	*sosa:Result*	*base:result/value_{value}_{unitOfMeasurement}*
Geometry	*geo:Geometry*	*base:geometry/{geometryType}_{geometryID}*
Census observation	*dwc:Occurrence*	*base:census/on_{censusDate}_{sesCategory}_{amountObserved}*
Event associated with observation	*bigeonto:BioEvent*	*base:bioevent/on_{censusDate}*
Location associated with observation	*dc:Location*	*base:geometry/point_{decimalLongitude}_{decimalLatitude}*
Observer	*foaf:person*	*base:person/{personID}*
SES category	*vocab:category*	*base:category/{categoryCode}*
Taxon	*dwc:Taxon*	*base:taxon/{mirounga_leonina}*
Related paper	*fabio:Expression*	*base:paper/{paperID}*

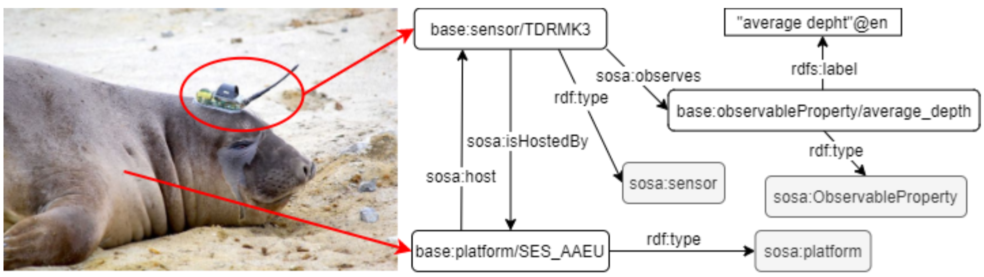

Fig. 1. SES Platform and TDR sensor description.

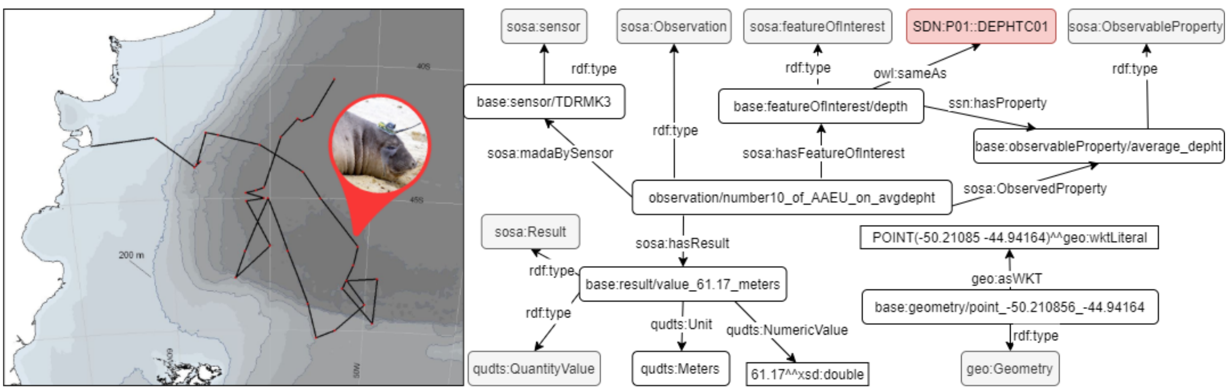

Fig. 2. Example of observation made by TDR on average depth in a particular location of the sea.

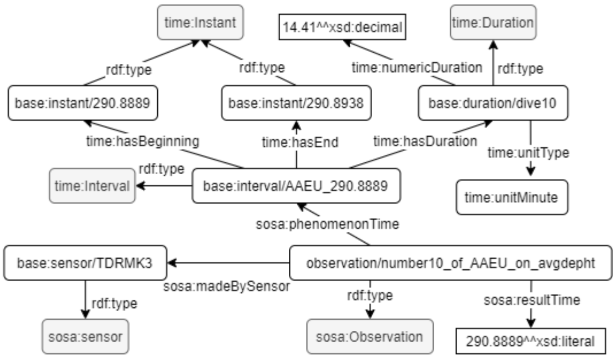

Fig. 3. Example of interval.

ganism belongs to a specific taxon (*dwc:Taxon*) whose scientific name is *Mirounga leonina*.

4.1.5. Publication

Each publication is represented as an instance of *fabio:Expression*. These expressions are also split by type of publication and thus using the respective subclasses of *fabio:Expression*. For instance, published books are represented with the class *fabio:Book* and so on. The shared properties for all the publications include: date of publication (*prism:publicationDate*), title (*dc:title*), doi (*prism:doi*), authors (*dc:creator*), abstract (*dc:abstract*) and file format (*dc:FileFormat*). Finally, documents reference, (*dc:references*), to the respective platforms (*sosa:platforms*), which have been involved in the results reported in those documents.

5. Data Transformation Process

To create Linked Open Data, a conversion needs to take place from the data contained in SES database into RDF. As explained in section 2, measurements produced by sensors, and census data are stored in MySQL server. Fields that are no longer used or that contain confidential data are excluded, for example data that is still being processed. Transformation process is done by D2RQ Platform [18], which consists of: The *D2RQ Mapping Language*, used to write mappings between database tables and RDF vocabularies or OWL ontologies. The *D2RQ Engine*, a SPARQL-to-SQL rewriter that can evaluate SPARQL queries over your mapped database, and *D2R Server*, a web application that provides access to the database via the SPARQL protocol, as Linked Data, and via a simple HTML interface. Listing 1 shows an example of mapping using the D2R mapping language, the data is extracted from the table whose name is *a0000_individuo_detalle*, which will be mapped as instances of *sosa:Platform*. The URI pattern is formed by the string *platform/SES_* and the primary key *ClaveU* between @@. The *urlify* function converts white space to (_) to form a valid URI.

Listing 1: Mapping example.

```
map:a0000_individuo_detalle a d2rq:ClassMap;
d2rq:dataStorage map:database;
d2rq:uriPattern "platform/SES_
            @@a0000_individuo_detalle.claveU
            |urlify@@";
d2rq:class sosa:Platform;
```

D2RQ runs in back-end at *http://linkeddata.cenpat-conicet.gob.ar* to browse structured data, it also has a SPARQL endpoint to be accessed from other applications, and a SPARQL explorer to query our own database in a friendly manner. One of the advantages that D2RQ provides us is that after mapping, if the database is updated, it is not necessary to rewrite the mapping. Key statistic presented in Table 5 was computed in March 2020.

5.1. Interlinking

The external links were generated manually, using a MySQL table created for this purpose, which is then mapped using D2RQ. This table has in one column the URI of our dataset, for example http://linkeddata.cenpat-conicet.gob.ar/resource/person/MZA and in another column the equivalent URI in an external dataset such as https://orcid.org/0000-0001-8851-8602, in this case they are the same person. See Figure 4.

When possible, in the case of publications, the instances of *fabio:Expression* class were linked to

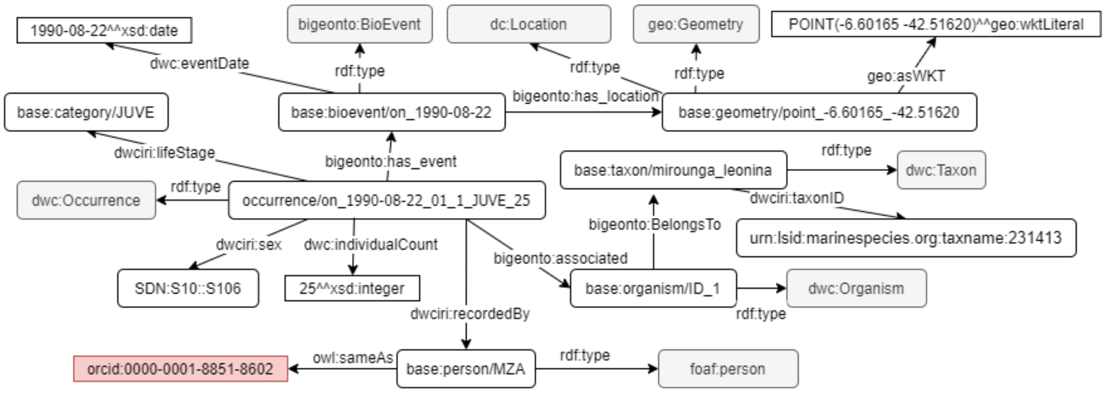

Fig. 4. Representation of an observation (*dwc:occurrence*) made during a census. The occurrence of an SES is performed by a person at a location and time.

Table 5
Key Statistics

Category	Resource
Total Nr. of Triples	16.8M
Nr. of classes	16
Nr. of properties	63
Nr. of platforms	9
Nr. of sensors	3
Nr. of observations	579.715
Nr. links to DBPedia	1
Nr. links to Wikidata	1
Nr. links to GeoNames	10
Nr. links to NERC	2
Nr. links to ORCID	7
Nr. links to OpenCitations	13

OpenCitation dataset [19], as show the Figure X. The relationship between these URIs is done using the property *owl:sameAs*.

Instances of *sosa:featureOfInterest* were linked to their corresponding URI in the NERC dataset, for example depth it has its equivalent URI, whose identifier is SDN:P01::DEPTHC01. Figure 2 shows the generated link in detail.

For external links referring to places, we use Geonames[8] a geographical database that contains over 11.8 million geographical names. The structure behind the data is the *Geonames ontology v3.2* [9], which closely resembles the flat-file structure. An individual in the database is an instance of type Feature and has a Feature Class (administrative divisions, populated places, etc.), a Feature Code (subcategories of Feature Class) along with latitude, longitude, etc. associated with it. In our case, the occurrence places such as beaches (instances of *geo:Feature* and *dc:Location*) where occurrences are recorded where linked to Geonames whenever possible.

To link instances of people *foaf:person* we use links to Open Researcher and Contributor Identifiers (ORCiDs) [20], They are intended to uniquely identify researchers so that those individuals can be correctly credited for their research work and links can be provided to express their professional affiliations. To date (April 2020) 26 external links have been created to different datasets.

5.2. *Dataset availability*

The SES dataset can be downloaded, navigated and queried using a SPARQL endpoint, and they are published under Creative Commons Attribution 4.0 International (CC BY 4.0)[10] License. All the criteria for five star Linked Data as defined in [21] are met. There is a description of the data online, the data is available in RDF, there are many links to structured vocabularies and metadata about the collection is made available. Our dataset characteristics are listed in Table 6.

6. Use Case: Accessing and analyzing data from dives and censuses

One crucial aspect is how to access and analyze data, and especially how to get only that part of data which is of interest for a given research question.

[8] https://www.geonames.org/
[9] http://www.geonames.org/ontology/documentation.html
[10] https://creativecommons.org/licenses/by/4.0/deed.en

Table 6
Technical details

URL	http://linkeddata.cenpat-conicet.gob.ar
SPARQL	http://linkeddata.cenpat-conicet.gob.ar/sparql
SNORQL	http://linkeddata.cenpat-conicet.gob.ar/snorql
Dump Data	10.17632/5nv5c7575w.2
VoID	http://linkeddata.cenpat-conicet.gob.ar/dataset
Licencing	CC BY 4.0

solves the access part, and SPARQL allows to query only a subset of the data.

To show the exploitation of the dataset, we developed a dashboard (link: https://cesimar.shinyapps.io/DiveAnalysisDashboard/) that allows you to consult the statistics of the dives and the routes taken. We use the R *flexdashboard*[11] package that allows generating web pages based on an R Markdown[12] document. To query our endpoint we use the SPARQL Package[13] allows you to directly import results of SPARQL SELECT queries into the statistical environment of R as a data frame. The following describes each module of the dashboard.

Diving statistics: This module summarizes the diving statistics (maximum depths recorded, number of dives, maximum temperatures and number of platforms). The information for each of the sensors used is also detailed. For bar charts, ggplot[14] library was used.

Dive Analysis: this module allows you to see by platform the most important variables registered during dives. Temperatures and depths, as well as duration can be displayed. The line chart was built using the plot_ly[15] library.

Platforms trips: This module retrieve the trips made by each platform and displays them on a map generated with the leaflet library[16], a filter can be made for each one if necessary. A spatial cluster analysis using the dbscan[17] algorithm is also provided to understand the distribution of SES at sea. The parameters can be configured by the user for their best adjustment.

Census statistics: This module allows analyzing the data of the censuses carried out during 1990 to 2017. Two charts were developed with ggplot, the first shows the annual population of SES grouped by category, while the second shows the trend of the SES breeding population. Figure 5 shows the main screen of the developed dashboard.

7. Discussion

This paper presents the publication as LOD of a biological and physical dataset, collected for more than 20 years, and stored with early objective of to study the environment influence on the foraging, reproductive performance and population trend of the SES initially available for a small research group to a global community. Our development with Linked Open Data improves the discoverability of the content of the database and could be applied at new knowledge-building and cross-disciplinary. For example, we expect the hydrographic profiles become a useful tool together physical samples resulting from other science programs, to assess ocean changes associated with the climate change.

The dataset comes from PV a geographic region under conservation regulations by UNESCO and there is a continue demand of the governmental authorities to develop spatial planning. This requirement helps the sustainability of the database, because it needs a high level of accuracy for data actualized of SES and access at other databases in a user-friendly manner. Coastal management Planning and Marine Spatial Planning (MSP) are concerned with the management of the distribution of human activities in space and time in and around seas and oceans to achieve ecological, economic and societal objectives and outcomes [22]. The next steps will be to promote the use vocabulary terms for discovery databases purposes of the institute CESIMAR, to allow the availability and suitability of data, to be used at regular review cycles of the MSP process. In addition, it would be desirable to access to the physical dataset collected by tourist and commercial vessels that overlap the same range in the southwest Atlantic Ocean. These hydrographic profiles could cover changes of the environment in all influence area of the SES distribution.

Acknowledgements

This work could not have been possible if it were not for the immense task of digitizing and ordering the data carried out by the Computer Scientist Maria Rosa Marin (recently retired). We grateful to PhD. Claudio

[11]https://rmarkdown.rstudio.com/flexdashboard/
[12]https://rmarkdown.rstudio.com/
[13]http://cran.r-project.org/web/packages/SPARQL/index.html
[14]https://cran.r-project.org/web/packages/ggplot2/index.html
[15]https://cran.r-project.org/package=plotly
[16]https://cran.r-project.org/web/packages/leaflet/leaflet.pdf
[17]https://cran.r-project.org/package=dbscan

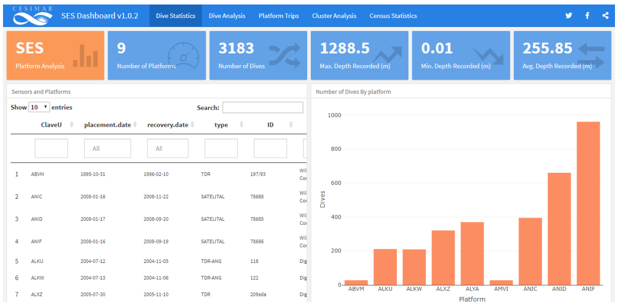

Fig. 5. Dashboard main screen developed for the analysis of diving and census data..

Campagna founder of the program for his vision of long lasting studies that allowed to keep the collection of records.

References

[1] T. Heath and C. Bizer, Linked data: Evolving the web into a global data space, *Synthesis lectures on the semantic web: theory and technology* **1**(1) (2011), 1–136.

[2] T. Tanhua, S. Pouliquen, J. Hausman, K. O'Brien, P. Bricher, T. de Bruin, J.J.H. Buck, E.F. Burger, T. Carval, K.S. Casey, S. Diggs, A. Giorgetti, H. Glaves, V. Harscoat, D. Kinkade, J.H. Muelbert, A. Novellino, B. Pfeil, P.L. Pulsifer, A. Van de Putte, E. Robinson, D. Schaap, A. Smirnov, N. Smith, D. Snowden, T. Spears, S. Stall, M. Tacoma, P. Thijsse, S. Tronstad, T. Vandenberghe, M. Wengren, L. Wyborn and Z. Zhao, Ocean FAIR Data Services, *Frontiers in Marine Science* **6** (2019), 440. doi:10.3389/fmars.2019.00440.

[3] Semantic Sensor Network Ontology W3C Editor's Draft 24 January 2020, 2020, [Online; accessed 27-Jan-2020].

[4] M. Zárate, G. Braun, P.R. Fillottrani, C. Delrieux and M. Lewis, BiGe-Onto: An Ontology-Based System for Managing Biodiversity and Biogeography Data, *Applied Ontology journal* (2020), 1–27–.

[5] M.N. Lewis and C. Campagna, Los elefantes marinos de Península Valdés, *Ciencia Hoy* **1** (2014), 187–194.

[6] UNESCO, Península Valdés World Heritage, 1999, [Online; accessed 17-Apr-2020].

[7] M. Lewis, C. Campagna and F. Quintana, Site fidelity and dispersion of southern elephant seals from Patagonia, *Marine Mammal Science* **12**(1) (1996), 138–147.

[8] M. LEWIS, C. CAMPAGNA and J. ZAVATTI, Annual cycle and inter-annual variation in the haul-out pattern of an increasing southern elephant seal colony, *Antarctic Science* **16**(3) (2004), 219–226–. doi:10.1017/S0954102004002020.

[9] C. Rutz and G.C. Hays, New frontiers in biologging science, *Biology Letters* **5**(3) (2009), 289–292. doi:10.1098/rsbl.2009.0089. http://rsbl.royalsocietypublishing.org/content/5/3/289.

[10] C. Campagna, A.R. Piola, M.R. Marin, M. Lewis and T. Fernández, Southern elephant seal trajectories, fronts and eddies in the Brazil/Malvinas Confluence, *Deep Sea Research Part I: Oceanographic Research Papers* **53**(12) (2006), 1907–1924.

[11] M. Compton, P. Barnaghi, L. Bermudez, R. GarcíA-Castro, O. Corcho, S. Cox, J. Graybeal, M. Hauswirth, C. Henson, A. Herzog et al., The SSN ontology of the W3C semantic sensor network incubator group, *Journal of Web Semantics* **17** (2012), 25–32.

[12] J. Wieczorek, D. Bloom, R. Guralnick, S. Blum, M. Döring, R. Giovanni, T. Robertson and D. Vieglais, Darwin core: An evolving community-developed biodiversity data standard, *PLoS ONE* (2012). ISBN 1932-6203 (Linking). doi:10.1371/journal.pone.0029715.

[13] H. de Koning, N. Rouquette, R. Burkhart, H. Espinoza and L. Lefort, Library for Quantity Kinds and Units: schema, based on QUDV model OMG SysML (TM), Version 1.2, Technical Report, CSIRO, Tech. Rep., 2011, retrieved: 04, 2016.[Online]. Available: http://www . . . , 2009.

[14] M. Perry and J. Herring, OGC GeoSPARQL-A geographic query language for RDF data, *OGC implementation standard* (2012).

[15] Time Ontology in OWL W3C Recommendation 19 October 2017, 2017, [Online; accessed 27-Jan-2020].

[16] S. Peroni and D. Shotton, FaBiO and CiTO: Ontologies for describing bibliographic resources and citations, *Journal of Web Semantics* **17** (2012), 33–43. doi:https://doi.org/10.1016/j.websem.2012.08.001. http://www.sciencedirect.com/science/article/pii/S1570826812000790.

[17] A. Leadbetter, R. Lowry and D. Clements, The NERC Vocabulary Server: Version 2.0, in: *Geophysical Research Abstracts*, Vol. 14, 2012.

[18] C. Bizer and R. Cyganiak, D2r server-publishing relational databases on the semantic web, in: *Poster at the 5th international semantic web conference*, Vol. 175, 2006.

[19] S. Peroni and D. Shotton, OpenCitations, an infrastructure organization for open scholarship, *Quantitative Science Studies* **1**(1) (2020), 428–444.

[20] L.L. Haak, M. Fenner, L. Paglione, E. Pentz and H. Ratner, ORCID: a system to uniquely identify researchers, *Learned Publishing* **25**(4) (2012), 259–264.

[21] K. Janowicz, P. Hitzler, B. Adams, D. Kolas, I. Vardeman et al., Five stars of linked data vocabulary use, *Semantic Web* **5**(3) (2014), 173–176.

[22] C. Ehler and F. Douvere, Marine Spatial Planning: a step-by-step approach toward ecosystem-based management. Intergovernmental Oceanographic Commission and Man and the Biosphere Programme. (2009).