

Towards a Global Record of Stocks and Fisheries

Yannis Tzitzikas^{1,2}, Yannis Marketakis¹, Nikos Minadakis¹, Michalis Mountantonakis^{1,2}, Leonardo Candela³, Francesco Mangiacrapa³, Pasquale Pagano³, Costantino Perciante³, Donatella Castelli³, Marc Taconet⁴, Aureliano Gentile⁴, Giulia Gorelli⁴

¹Institute of Computer Science, FORTH-ICS, Heraklion, Greece, e-mail: {tzitzik, marketak, minadakn, mountant}@ics.forth.gr

²Computer Science Department, University of Crete, Heraklion, Greece

³Consiglio Nazionale delle Ricerche, Pisa, Italy, e-mail: {leonardo.candela,

francesco.mangiacrapa, pasquale.pagano, costantino.perciante, donatella.castelli}@isti.cnr.it

⁴Food and Agriculture Organization of the United Nations, Rome Italy, e-mail: {marc.taconet, aureliano.gentile, giulia.gorelli}@fao.org

Abstract. The collation of information for the monitoring of fish stocks and fisheries is a difficult and time-consuming task, as the information is scattered across different databases and is modelled using different formats and semantics. Our purpose is to offer a unified view of the existing stocks and fisheries information harvested from three different database sources (FIRMS, RAM and FishSource), by relying on innovative data integration and manipulation facilities. In this paper, we describe the activities carried out to realize the Global Record of Stocks and Fisheries (GRSF) which aims at offering an integrated and enriched view on data about fish stocks and fisheries from the database sources. More specifically we describe the model, the workflow and the software components for producing GRSF records and make them easily available to the users.

Keywords: fish stock, fishery, semantic data integration, data publication

1 Introduction

Fish Stocks are groups of individuals of a species occupying a well-defined spatial range independent of other stocks of the same species, e.g. swordfish in the Mediterranean Sea¹. A *Fishery* is a unit determined by an authority or other entity that is engaged in raising and/or harvesting fish. Typically, the unit is defined in terms of some or all of the following: people involved, species or type of fish, area of water or seabed, method of fishing, class of boats and purpose of activity, e.g. Fishery for Atlantic cod in the area of East and South Greenland². Information about Fish Stocks and Fisheries is widely used for the monitoring of their status, and to

¹ <http://firms.fao.org/firms/resource/10025/en>

² https://www.fishsource.org/stock_page/688

Copyright © 2017 for this paper by its authors. Copying permitted for private and academic purposes.

Proceedings of the 8th International Conference on Information and Communication Technologies in Agriculture, Food and Environment (HAICTA 2017), Chania, Greece, 21-24 September, 2017.

identify appropriate management actions (Hilborn & Walters, 2013), with the ultimate goal of sustainable exploitation of marine resources. For these reasons completeness, adequacy and validity of information is crucial. Although this key role, there is no “one stop shop” for accessing stocks and fisheries data. Such information is usually collected (and produced as a result of data analysis) by the fishery management authorities at regional, national and local level. Therefore, the overall information is scattered across several databases, with no standard structure due to the specific local needs of the different bodies. Furthermore, the guidelines for populating existing registries are therefore heterogeneous, and every registry is actually a “database silo” that is not expected to interoperate with others to offer a global view on existing information.

Our objective is to construct a Global Record of Stocks and Fisheries (for short GRSF) capable of containing the corresponding information categorized into uniquely and globally identifiable records. Instead of creating yet another registry, we focus on producing GRSF records by using existing data. This approach does not invalidate the process being followed so far, in the sense that the organizations that maintain the original data are expected to continue to play their key role in collecting and exposing them. In fact, GRSF does not generate new data, rather it collates information coming from the different database sources, facilitating the discovery of inventoried stocks and fisheries arranged into distinct domains.

The advantages of this approach include: (a) offering increased data coverage compared to the single sources of information, (b) integrating information and unique identification of stocks and fisheries coming from the different database sources, and (c) answering queries that would be impossible to be answered from the individual database sources. These characteristics meet the needs of the main business cases that are: (i) supporting the compilation of stock status summaries at regional and global level and (ii) providing services for the traceability of sea-food products.

In this paper we introduce the process that has been used for constructing and easily maintaining GRSF. In fact, GRSF maintenance is an almost continuous activity since data providers can constantly offer new or revised information. The rest of the paper is organized as follows: Section 2 discusses the motivation and the requirements. Section 3 describes the architecture and the technical components for realizing GRSF. Section 4 discusses the current results. Finally, Section 5 concludes and identifies directions for future work and research.

2 Motivation and Settings

The objective of GRSF is to act as a “one stop shop” for stocks and fisheries records. It realizes an innovative environment supporting the collaborative production and maintenance of a comprehensive and transparent global reference set of stocks and fisheries records, that will boost regional and global stocks and fisheries status and trend monitoring, as well as, responsible consumer practices. To this end, a selected set of data sources is exploited for delivering relevant information. To ensure a high quality final product, a set of guidelines and standards has been identified which are described later in this section.

2.1 The Data Sources

Below we describe the three database sources that have been used so far to harvest stocks and fisheries information. These sources are (a) Fisheries and Resources Monitoring System (FIRMS), (b) RAM Legacy Stock Assessment database, and (c) FishSource. The rationale for the selection of these sources, is that they contain complementary information (both conceptually and geographically). More specifically FIRMS is mostly reporting at regional level, while RAM is reporting at national or subnational level, and FishSource is more focused on the fishing activities. All of them contribute to the overall aim to build a comprehensive and transparent global reference set of stocks and fisheries records that will boost regional and global stocks and fisheries status and trend monitoring as well as responsible consumer practices. Since the construction of GRSF is an iterative process, we will support integrating contents from these three sources in early releases of GRSF, and in future we will investigate exploiting new ones (i.e. FAO Global Capture Production Statistics database³).

FIRMS (Fisheries and Resources Monitoring System)⁴ provides access to a wide range of high-quality information on the global monitoring and management of stocks and fisheries. It collects data from 14 intergovernmental organizations (that are partners of FIRMS) and contains information about the status of more than 600 stocks and 300 fisheries. The information provided by the organizations is ingested in a database and published in the form of XML backbone fact sheets.

RAM (RAM Legacy Stock Assessment Database)⁵ provides information exclusively on the fish stocks domain. It is a compilation of stock assessment results and time series of stock status indicators for commercially exploited marine populations from around the world. The assessments are assembled from 21 national and international management agencies for approximately one thousand stocks. RAM contents are stored in a relational database and are publicly available by releasing versions of the database in MS Access and Excel format.

FishSource⁶ compiles and summarizes publicly available scientific and technical information about the status of fish stocks and fisheries. It includes information about the health of stocks, the quality of their management, and the impact of fisheries on the rest of the ecosystem. It is mainly exploited from seafood industry for assisting in taking the appropriate actions for improving the sustainability of the purchased seafood. Information in FishSource is organized into fishery profiles associated with the exploited stocks, and currently contains more than 2,000 fishery profiles.

2.2 Requirements

The selected database sources were originally constructed to fulfil different requirements and needs. Furthermore, they have been developed and are maintained from different initiatives. As a result, they are using different standards, data models,

³ <http://www.fao.org/fishery/statistics/global-capture-production/en>

⁴ <http://firms.fao.org/firms/en>

⁵ <http://ramlegacy.org>

⁶ <http://www.fishsource.com/>

conceptualizations and terminologies for capturing similar information. As an example consider the fish species that are included in a particular stock or fishery; they can be identified either using (a) their scientific name (e.g. *Thunnus albacares*), (b) their common name in any language (e.g. Yellowfin tuna in English), or (c) standard codes for identifying them (e.g. YFT⁷). Furthermore, the different data sources use diverse criteria for identifying the uniqueness of a stock or fishery, as well as diverse conventions for naming their records.

GRSF aims at harmonizing the harvested information by adopting a set of standards that have been discussed and agreed with representatives of the database sources. In particular, these standards have been identified by two technical working group meetings that have been organized. The working groups have defined which are the international standards that will be used (e.g. FAO 3Alpha codes for species, ISO3 country codes for flag states), which values define the uniqueness of a stock or a fishery record, which values are mandatory to accept a record as a complete one, as well as guidelines for generating unique and global identifiers (both human and machine interpretable) and names for the GRSF records. A detailed description of a GRSF record with respect to those guidelines can be found in Section 2.3.

The main challenge for the construction of the GRSF is the ability to semantically integrate data coming from different data sources. To tackle this challenge, we decided to rely on semantic web technologies and use top level ontologies. The best candidate is the MarineTLO (Tzitzikas et al., 2016-a) which provides (a) consistent abstractions or specifications of concepts included in all data models or ontologies of marine data sources and (b) the necessary properties to make GRSF a coherent source of facts relating observational data with the respective spatiotemporal context and categorical domain knowledge. The rationale is that we map attributes from different data sources into classes and properties of the top-level ontologies. To this end we could also mention works like (Pham et al., 2016) that automate the mapping process using machine-learning techniques.

2.3 The GRSF record

Each GRSF record is composed of several fields to accommodate the incoming information and data. The fields can be functionally divided into time-independent and time-dependent. The first group contains the identification, descriptive and other information which describes various aspects of a stock or fishery, and the latter contains the stocks and fishery indicators. In general, there are two types of GRSF records: (a) stocks and (b) fishery GRSF records. Both types of records share some common metadata like their time-independent information. Furthermore, records are assigned information about areas and their original sources. Finally, each record is assigned several time-dependent information modeled as dimensions. In the case of stock GRSF records, the dimensions refer to abundance levels and exploitation rates. In the cases of fishery GRSF records, the dimensions refer to catches and landings indicators. We could say that a GRSF record resembles a data item in a database and as such we are describing its corresponding details in the schema shown in Fig. 1.

⁷ According to FAO 3Alpha code <http://www.fao.org/fishery/collection/asfis/en>

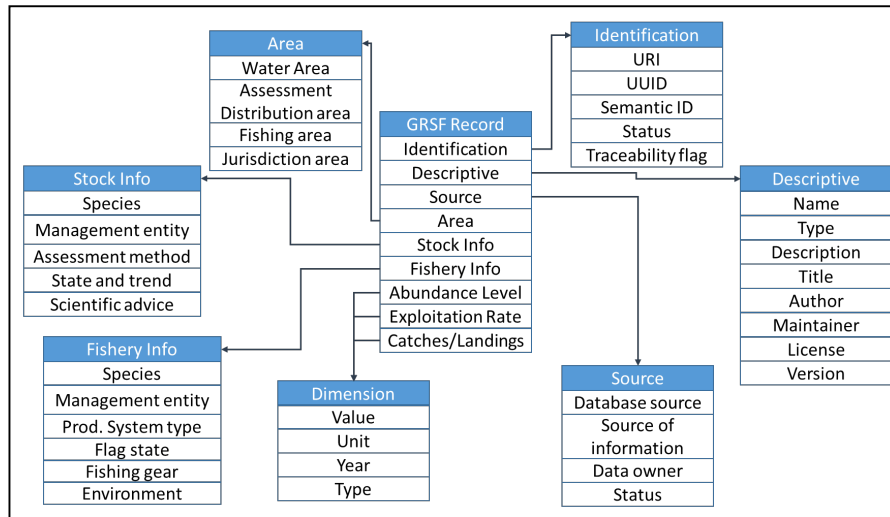


Fig. 1 The STAR schema of a GRSF record

2.4 The Process

The process for constructing GRSF consists of a sequence of steps which are shown in Fig. 2. Below we describe these steps in detail. The technical components that carry out each step of the process are described in detail in Section 3.

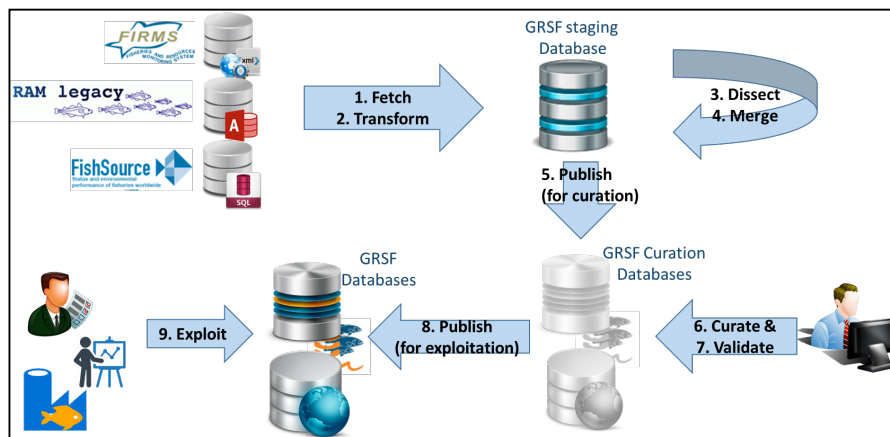


Fig. 2 The steps required for constructing and exploiting GRSF

Fetch. GRSF does not affect the data from the remote database sources. This means that the maintainers of the database sources will continue to update them in their own systems. For including the providers' data in the GRSF it is important to periodically fetch the raw data (in their original form) or the data in a different format or view if they are exposed using particular services (i.e. in other formats like JSON

or XML). In particular FIRMS offers a set of services that exposes their contents in XML format, RAM publish their MS Access database in their website, and FishSource exposes specific parts of their relational database as JSON data through a set of web services.

Transform. After fetching the data it is important to transform them so that they have a similar structure and semantics. At this stage data is transformed from XML, JSON and MS Access to RDF format. Specifically, data is transformed into instances of the MarineTLO ontology with respect to the identified GRSF requirements. Information harvested from the database sources will be mapped to the agreed GRSF standards, when not already compliant. Furthermore, during this step a set of proximity rules are applied (using the species, area and gear fields) for identifying similar records. This creates groupings of similar records that are being used in subsequent phases (during the curation & validation phase).

Dissect. This step is important for complying with the standards, for traceability aspects. In some cases, sources contain aggregated information in their records. For example, in a single fishery record there could be included more than one species, fishing gears or flag states. These aggregated records are therefore dissected to produce new GRSF records, each containing one single value for the above-mentioned fields, and thus complying with the requirements for traceability.

Merge. This step ensures that the contents that have been added in the GRSF staging database are properly connected based on a set of criteria. This is achieved by linking records that have the same values on particular fields (specifically time-independent values) for producing a new single GRSF record. For example, if there are stock records having the same species and water area, we can merge them into a single stock. During this process, we also use external knowledge to detect similarities among different names and terminologies used in the database sources (i.e. species names). The time-dependent information for the merged records will be kept distinct although collated and associated to the final merged GRSF record, with clear indication of the database source and the reference year.

Publish (for curation). The contents of the GRSF staging database are being replicated into a public GRSF database, which is actually a triple-store. The triple-store can be used as a reference endpoint for answering complex queries about stocks and fisheries records. Furthermore the contents are published in a data catalogue offered through the D4Science (Candela et al., 2014) infrastructure. These resources allow the experts inspecting the contents of the GRSF and curate them appropriately. During this step, Universally Unique Identifiers (UUID) and human readable semantic identifiers are generated and associated to each GRSF record. The former are generated based on a standard algorithm and are used to uniquely identify records. The latter are generated using various GRSF fields and populated with standard codes and allow the identification and interpretation of records by humans.

Curate & Validate. During this step, a community of experts browse over the GRSF records and curate them in various ways. At this stage, the GRSF records are in a pending status waiting for approval by a human expert. During this process, the experts are able to either approve or reject a record, as well as to suggest alternative processes for merging records and to attach annotations with a narrative text.

Publish (for exploitation). The GRSF records that has been approved during the previous phase are being published into public and read-only databases as final GRSF products that can be exploited from the communities of interest.

3 Software Components and Architecture

The D4Science infrastructure and gCube technology (Assante et al., 2016) enable the development of Virtual Research Environments (VREs) that provide the users with a web-based set of facilities to accomplish various tasks. For the purpose of GRSF, we developed the appropriate VREs acting as a gateway for the “one stop shop” for stocks and fisheries records. More specifically we exploit the data cataloguing facilities of the infrastructure for manipulating and exposing GRSF records to the wide audience.

The core component for constructing GRSF is MatWare (Tzitzikas et al., 2014). MatWare is a framework that automates the process of constructing semantic warehouses. By using the term semantic warehouse we refer to a read-only set of RDF triples fetched and transformed from different sources that aims at serving a particular set of query requirements. MatWare automatically fetches contents from the underlying sources using several access methods (e.g. SPARQL endpoints, HTTP accessible files, JDBC connections, several file format transformers). The fetched data are transformed into RDF descriptions using appropriate mappings (Marketakis et al., 2016), and stored in a RDF triplestore supporting several levels of description for preserving provenance information. One of its distinctive features, is that it allows evaluating the connectivity of the semantic warehouse. Connectivity refers to the degree up to which the contents of the semantic warehouse form a connected graph that can serve ideally in a correct and complete way the query requirements, while making evident how each source contributes by using a set of connectivity metrics. MatWare is a fully configurable tool and can be easily extended using plugins. For the purposes of GRSF we have extended it with plugins for fetching and transforming the data from their original formats, plugins for supporting the merging and dissection steps, as well plugins for publishing the data into the catalogue supporting both the curation and validation phase, as well as the consumption phase.

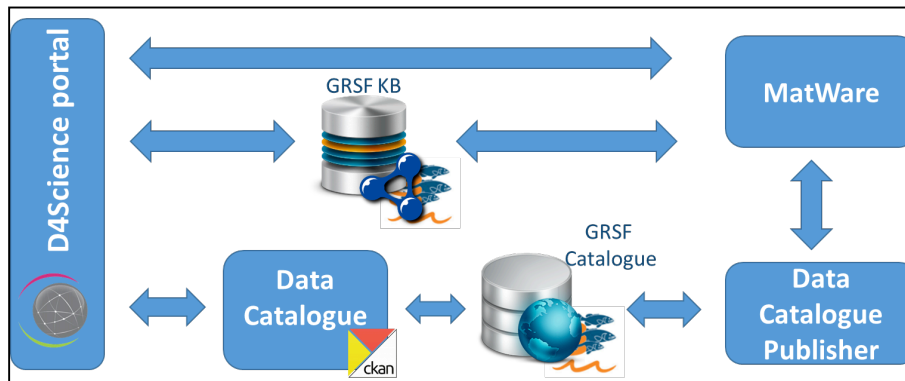


Fig. 3 The GRSF construction deployment setting

Fig. 3 shows the overall technical deployment for the construction and maintenance of the GRSF. MatWare is responsible for the activities that construct the GRSF (as they are described in Section 2.4) and publishing them in the GRSF Knowledge Base (GRSF KB) and in the GRSF Catalogue. For the latter it exploits the component Data Catalogue publisher which carries out the necessary activities for ingesting GRSF records into the CKAN-based Catalogue instance offered by the D4Science infrastructure. Finally all the above components are controlled and interacted through the D4Science portal facilities of the GRSF VREs.

4 Evaluation – Discussion

In order to assist the experts during the process of inspecting the GRSF records it has been decided to keep and publish both the initial records from the database sources as well as the final GRSF ones. For each one of the final records we preserve the provenance information about the initial records it has been derived from. This will allow the experts to quickly identify the problematic sources in cases of erroneous final records that have to be rejected and undergo a different handling. To distinguish the initial records from the final ones we used the notion of named graphs in the GRSF KB; the resources coming from each of the initial data sources have been added in a particular named graph, and the final records on a different one – preserving however the links (in terms of URIs) to the corresponding initial resources that exist in different named graphs. A similar approach has been carried out in the GRSF data catalogue, where the records are distinguished using groups (e.g. FIRMS Stock, RAM Stock, GRSF Stock, GRSF Fishery, etc.).

Table 1. Summary of the information fetched and integrated into GRSF

	FIRMS	RAM	FishSource	GRSF
Stock Records	491	989	873	2,187
Fishery Records	190	-	2,203	7,486
Species	578	264	389	1,204
Water Areas	264	636	302	1,181
Fishing Gears	40	-	59	97
Flag States	69	-	96	163
Assessment Methods	58	74	-	110
Scientific Advices	243	-	326	506

Table 1 summarizes some statistics about the results of the first version of GRSF. We should note here that only a limited number of merging activities took place. However, we foresee that in future versions of GRSF we will fully support merging activities. The table contains the total number of Stock and Fishery records, as well as the distinct number of particular information like for example the distinct number of species, water areas and others. What is important to describe here is the high number of the fishery records in the GRSF database (7,486 records) compared to the summary of the fishery records in the initial sources (2,393 records in total). This is due to the merging and dissection processes. More specifically many fishery records (especially from FishSource) had multiple values for their: (a) target species, (b) fishing gears, and (c) flag states. These records have been dissected to contain single values for these fields, a decision that was taken for being compliant with the guidelines of GRSF to meet the traceability business case, as agreed between the representatives of the database sources.

The merging process implies the collation of information, thus filling gaps of knowledge that may occur in the single database sources. In addition integrating the data sources into a semantic warehouse allows us to create a knowledge graph that interconnects all the relevant information following the Linked Data principles (Heath & Bizer, 2011). For instance the three original database sources use different ways for identifying the targeted species; FIRMS use their common name in English (e.g. yellowfin tuna), RAM use their scientific Latin name (e.g. *Thunnus albacares*) and FishSource use their FAO 3-Alpha code (e.g. YFT). This information is interconnected in GRSF using the appropriate properties of the top level classes as shown in Fig. 4. The apparent advantage is that users exploiting GRSF will be able to find records using any of the above names.

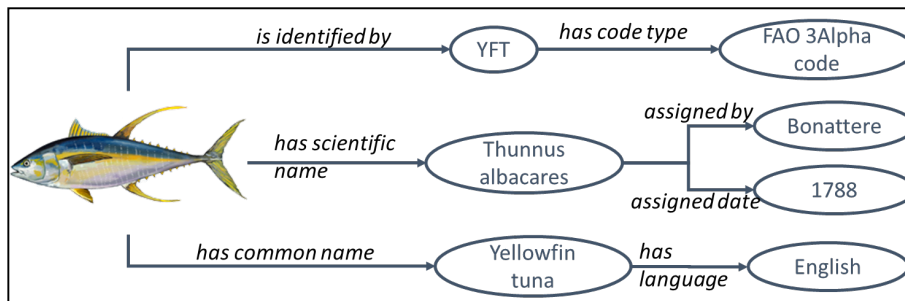


Fig. 4 Different ways for identifying a fish species

The semantic warehouse enables querying data from all the underlying sources in a uniform manner. Table 2 shows the results of the connectivity metrics (Mountantonakis et al., 2016) (common URIs, literals and triples). GRSF – as a source – scores the highest value, which justifies that it contains highly connected and valuable information.

Table 2. Connectivity metrics for GRSF sources as computed by MatWare

	FIRMS	Common URIs			GRSF	FIRMS	Common Literals			GRSF	Unique Triples (%)	Value
		RAM	Fish Source	GRSF			RAM	Fish Source	GRSF			
FIRMS	1.0	0.16 %	0.19 %	50.49 %	1.0	3.19 %	7.85 %	45.17 %	82.58 %	0.1087		
RAM		1.0	0.14 %	41.71 %		1.0	4.41 %	27.41 %	93.07 %	0.1038		
FishSource			1.0	18.24 %			1.0	8.07 %	97.31 %	0.2011		
GRSF				1.0				1.0	92.28 %	0.3402		

The warehouse can also be exploited as a valuable source of entity names for improving the quality of automatic semantic annotation of texts and documents, e.g. by using tools like XLink (Fafalios et al., 2015) that can exploit external SPARQL endpoints. Moreover one could provide exploratory search services over its contents by adopting the related approaches that have been developed for RDF datasets (see Tzitzikas et al., 2016-b for a survey) or by first defining semantic views and then exploring them through Hippalus, just like it has been done for fish species (Tzitzikas et al., 2016-c). Finally, it can be exploited for semantically annotating search results coming from external search systems (Fafalios et al., 2014).

The contents of the GRSF are currently exposed through the D4Science portal. More specifically two Virtual Research Environments have been created for sharing knowledge about Stocks and Fisheries, and supports the GRSF business cases. The first one is for the community of experts for carrying out the curation and validation activities, and another one is for public use. Fig. 5 shows some indicative screenshots from the corresponding VREs. By using it, it is possible to search for records by keyword based search as well as faceting by tags (e.g. commercial species, fishing area, fishing typology) and groups (e.g. record type, source provider). For each record the catalogue offers a user-friendly view of its content and cater for visualizing the time-dependent information associated with it, as well as any other multimedia resource attached to the record.

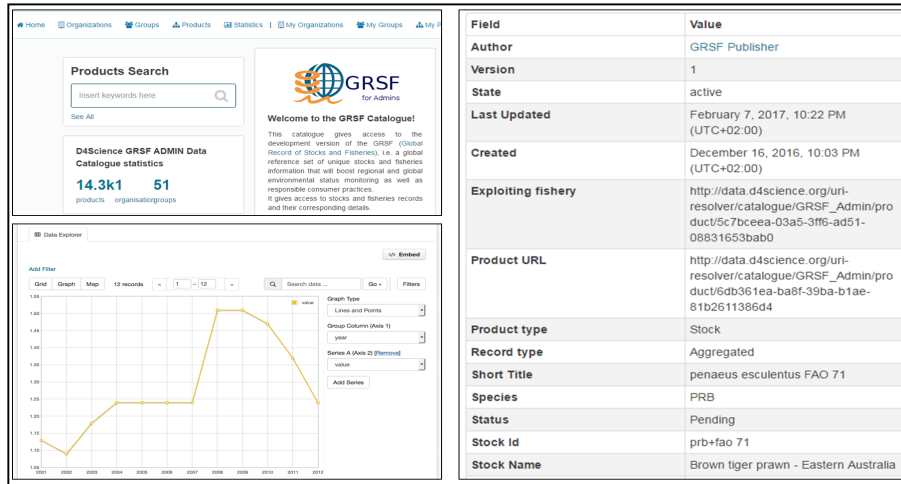


Fig. 5 Exposing GRSF through the GRSF Virtual Research Environment

As regards the GRSF construction and maintenance we could say that it consists of three different phases: the design, the implementation and the maintenance phase. During the design phase the activities for defining the mappings from the original sources to the targeted format, and the merging and dissection rules are defined. These processes require human effort and especially they require close collaboration and interactions with the organizations maintaining the original sources, to make sure that the semantics of the original data are modelled adequately with respect to the target models. The output of the design phase is a set of mappings and rules formally expressed, so that can be used during the implementation phase. The implementation phase is an automated process that exploits the results of the design phase and realizes the processes described in Section 2.4. As regards the maintainability of the GRSF, it is an automated process that can be triggered whenever new data exist in the original data sources, assuming of course that their structure has been preserved, otherwise the mappings should also be updated. Finally, we should mention that during the maintenance phase, any changes made to records by the users (i.e. changes in the status of a record) are being preserved.

For the particular setting of GRSF, the implementation phase took approximately 47 hours to be completed, and consists of the time for fetching and transforming the data (~45 minutes), merging and dissecting them (~ 30 minutes) and publishing them sequentially in the D4Science Data catalogue (~ 45 hours). In future we plan to parallelize the last subtask for being able to publish multiple records simultaneously.

5 Conclusion – Future Work

In this paper we introduced a process for providing a unified view of several stocks and fisheries databases, by relying on semantic web technologies and innovative hybrid data infrastructures. The resulting Global Record of Stocks and

Fisheries integrates data from three data sources, and contains more than 9,500 records about stocks and fisheries. It can be seen as a core knowledge base supporting the collaborative production and maintenance of a comprehensive and transparent global reference set of stocks and fisheries records. This is accomplished because of the processes that were applied during the construction, that guarantee the unique identification of stock and fisheries and the easy access to the information associated to a particular stock or fishery. In addition, during the validation step, the experts can validate the information of the GRSF records which also allows them spotting errors in their original sources, because their provenance is also preserved.

We plan to release newer versions of GRSF periodically that will fully support the merging activities. Apart from the new versions, we also plan to perform an inventory for more database sources to be included. Apart from the technical details of GRSF, there are several issues that are worth further work and research including (a) support of update operations for specific records in the GRSF data source or the original sources, (b) parallelization of the construction and maintenance phases for fastening them, (c) offer advanced discovery services based on spatio-temporal information, and (d) investigation of whether machine learning techniques could be exploited for automating or assisting the curation and validation of GRSF records.

Acknowledgments. This work has received funding from the European Union's Horizon 2020 research and innovation programme under the BlueBRIDGE project (Grant agreement No 675680).

References

1. Candela L., Castelli D., Manzi A. and Pagano P. (2014) Realising Virtual Research Environments by Hybrid Data Infrastructures: the D4Science Experience. International Symposium on Grids and Clouds (ISGC) Proceedings of Science PoS (ISGC2014).
2. Assante M., Candela L., Castelli D., Coro G., Lelii L. and Pagano P. (2016) Virtual research environments as-a-service by gCube. PeerJ Preprints 4:e2511v1 doi:10.7287/peerj.preprints.2511v1.
3. Tzitzikas Y., Minadakis N., Marketakis Y., Fafalios P., Allocca C., Mountantonakis M. and Zidianaki I. (2014) MatWare: Constructing and Exploiting Domain Specific Warehouses by Aggregating Semantic Data. In proc of the 11th Extended Semantic Web Conference (ESWC'14), Crete, Greece
4. Marketakis Y., Minadakis N., Kondylakis H., Konsolaki K., Samaritakis G., Theodoridou M., Flouris G. and Doerr M. (2016) X3ML Mapping Framework for Information Integration in Cultural Heritage and beyond. Int. Journal on Digital Libraries, pp 1-19. Springer. DOI 10.1007/s00799-016-0179-1.
5. Tzitzikas Y., Allocca C., Bekiari C., Marketakis Y., Fafalios P., Doerr M., Minadakis N., Patkos T. and Candela L. (2016-a) Unifying Heterogeneous and Distributed Information about Marine Species through the Top Level Ontology

MarineTLO. Emerald Group Publishing Limited 50(1)
<http://dx.doi.org/10.1108/PROG-10-2014-0072>.

6. Hilborn R. and Walters C. (2013) Quantitative fisheries stock assessment: Choice, Dynamics and Uncertainty. Springer Science & Business Media, ISBN 978-1-4615-3598-0.
7. Heath T. and Bizer C. (2011) Linked Data: Evolving the web into a global data space. Synthesis lectures on the semantic web: theory and technology. pp. 1-136.
8. Fafalios P., Baritakis M. and Tzitzikas Y. (2015) Exploiting Linked Data for Open and Configurable Named Entity Extraction. *International Journal on Artificial Intelligence Tools* 24(2)
9. Tzitzikas Y., Manolis N. and Papadakos P. (2016-b) Faceted Exploration of RDF/S Datasets: A Survey, *Journal of Intelligent Information Systems (JIIS)*.
10. Tzitzikas Y., Bailly N., Papadakos P., Minadakis N. and Nikitakis G. (2016-c) Using preference-enriched faceted search for species identification. *IJMSO* 11(3) pp. 165-179.
11. Fafalios P., Papadakos P. and Tzitzikas Y. (2014) Enriching Textual Search Results at Query Time Using Entity Mining, Linked Data and Link Analysis. *International Journal of Semantic Computing* 8(4) pp. 515-544
12. Pham M., Alse S., Knoblock C.A. and Szekely P. (2016) Semantic Labeling: A domain-independent approach. In proceedings of the 15th International Semantic Web Conference (ISWC 2016), Japan.
13. Mountantonakis M., Minadakis N., Marketakis Y., Fafalios P. and Tzitzikas Y. (2016). Quantifying the Connectivity of a Semantic Warehouse and Understanding its Evolution over Time. *International Journal on Semantic Web and Information Systems (IJSWIS)* 12(3), pp. 27-78, DOI: 10.4018/IJSWIS.2016070102