PREDICAT: a semantic service-oriented platform for data interoperability and linking in earth observation and disaster prediction

Maroua Masmoudi, Hela Taktak, Sana Ben Abdallah Ben Lamine, Khouloud Boukadi, Mohamed Hedi Karray, Hajer Baazaoui Zghal, Bernard Archimède, Michael Mrissa, Chirine Ghedira

PREDICAT: A semantic service-oriented platform for data interoperability and linking in earth observation and disaster prediction

Maroua Masmoudi[1,3], Hela Taktak[2,5], Sana Ben Abdallah Ben Lamine[1], khouloud Boukadi[2], Mohamed Hedi karray[3], Hajer Baazaoui Zghal[1], Bernard Archimede[3], Michael Mrissa[4], Chirine Ghedira Guegan[5]

[1]Riadi Laboratory, University of Manouba, Manouba, Tunisia
{sana.benabdallah,hajer.baazaouizghal}@riadi.rnu.tn

[2]Mir@cl Laboratory, Sfax University, FSEGS, Sfax Tunisia
{hela.taktak, khouloud.boukadi}@gmail.com

[3]LGP Laboratory, ENIT, Tarbes, France
{maroua.masmoudi, mkarray, bernard.archimede}@enit.fr

[4]InnoRenew CoE, Livade 6, 6310 Izola, Slovenia, FAMNIT, University of Primorska, Koper, Glagoljaška 8, 6000, Slovenia
michael.mrissa@innorenew.eu

[5]University of Lyon, CNRS, IAE University of Lyon 3, LIRIS, UMR5205 Lyon, France
chirine.ghedira-guegan@univ-lyon3.fr

Abstract— **The increasing volume of data generated by earth observation programs such as Copernicus, NOAA, and NASA Earth Data, is overwhelming. Although these programs are very costly, data usage remains limited due to lack of interoperability and data linking. In fact, multi-source and heterogeneous data exploitation could be significantly improved in different domains especially in the natural disaster prediction one. To deal with this issue, we introduce the PREDICAT project that aims at providing a semantic service-oriented platform to PREDIct natural CATastrophes. The PREDICAT platform considers (1) data access based on web service technology; (2) ontology-based interoperability for the environmental monitoring domain; (3) data integration and linking via big data techniques; (4) a prediction approach based on semantic machine learning mechanisms. The focus in this paper is to provide an overview of the PREDICAT platform architecture. A scenario explaining the operation of the platform is presented based on data provided by our collaborators, including the international intergovernmental Sahara and Sahel Observatory (OSS).**

Keywords-Earth observation; disaster prediction; data interoperability; data integration; ontology, service computing;

I. INTRODUCTION

In recent years, natural disasters have becoming more frequent and intense all around the world. The need for environmental monitoring has involved developments in information and Earth Observation (EO) programs such as the Copernicus program [5], the National Oceanic and Atmospheric Administration (NOAA) [21] and the Sahara and Sahel Observatory (OSS) [34]. EO programs are used to observe, monitor and assess the status of, and changes in, the environment. These systems are becoming increasingly important so that a vast amount of EO data is collected daily. These voluminous data could be significantly exploited in different domains. Examples include economical applications, environmental monitoring, phenomena understanding, decision-making during extreme weather crisis and disaster prediction. Disaster prediction is one of the most important application of data exploitation. Despite the availability of large amounts of data, the usage of EO data is still limited due to the lack of interoperability and data linking [13]. In fact, the overwhelming amount of data has worsened heterogeneity problems, as has the types of observation sources generating data in heterogeneous formats (databases, files and rasters) and heterogeneous semantics (synonymy, polysemy, etc...). For instance, in the sentence "Maps of daily temperature and precipitation are produced", an expert would recognize that the observation is "temperature" but would not be able to determine the details related to the temperature concept (atmospheric temperature, sea surface temperature, etc.).

Integrating this huge number of data known as big data is a real challenge since retrieving EO data from different sources involves the use of different APIs, if these latter exist. The issue of managing data derived from EO programs, in terms of access, pricing, data rights and other aspects, is commonly difficult. On the one hand, there are free of charge and open access data. On the other hand, data can be extracted only after agreeing to specific laws and regulations. Access to data requires to be user-friendly to reach common understanding and decision making for various prediction systems.

Currently, there are several ongoing projects[1][2][3] that are aiming to solve the integration problem. All these projects

[1] http://www.i-react.eu/
[2] http://beaware-project.eu/

similarly aim to semantically integrate heterogeneous data coming from big data sources such as sentinel data, including data provided by citizens through social media. However, data storage and management with traditional data management platforms is difficult [25]. As the number of data sources and the type of data stores augment, data access needs to be made easier for better real-time prediction

In this paper, we present the PREDICAT (PREDIct natural CATastrophes), that aims at providing a semantic service-oriented platform for data interoperability and linking in EO and disaster prediction. PREDICAT aims 1) to integrate EO data coming from several sources such as NOAA and OSS, including that provided by citizens, 2) to provide a decision support solution to analyze in real time all the useful data in order to effectively prevent and/or react against natural disasters, through semantic linking of information. The integration is performed at the semantic level, to ensure semantic interoperability between data and provide reasoning mechanisms; and at the service layer, by providing adequate services to access and extract data with any format or structure, in real-time, in order to guarantee faster data management. The PREDICAT platform also tackles data access and storage problems through service implementation, to hide the heterogeneity of data sources and allow interoperability between EO data systems. Such interoperability could help experts to detect possible disasters through the combination of pieces of knowledge coming from different sources.

Since data have poor semantics, the main objective of our platform is to have a global view of all data through semantic linking of information, and to produce warnings and real-time decisions to effectively prevent natural disasters. Furthermore, among other encountered research problems related to big data domain is the large-scale data exchange of heterogeneous datasets.

The rest of this paper is organized as follows. In Section 2, we give an overview of existing work on services and APIs for EO data (section 2.1) and ontology-based prediction systems (section 2.2). We also expose our motivations and the main goals of this research (section 2.3). Then in section 3, we present the PREDICAT platform architecture and its main components. In section 4, we provide a scenario example demonstrating the applicability of T for the integration and the management of data. Finally, we conclude and present our future work in section 5.

II. BACKGROUND AND MOTIVATIONS

This section presents the background related to EO APIs and identifies issues of their usage and access (Section 2.1). Besides, it presents related work on ontology-based prediction systems (Section 2.2).

A. Services and APIs for EO data

1) EO data: Usually EO data sets are described using metadata that include information related to the data such as its collection time, the author or the data source, the file size, etc. Initial works have used metadata expressed in natural language with plain-text [26] [29], which may involve ambiguous and inaccurate expressions and may cause low-speed processing problems. As an alternative, the meteorological and climate scien-tists group proposed to adopt the binary formats such as the GRidded Binary (GRIB), the Network Common Data Format (NetCDF) [32] [33], the Hierarchical Data For-mat (HDF) [3] and the BIL (Band Interleaved by Line) format for images relating to soils textures. However, exploiting these data sources is facing multiple challenges, among them heterogeneity of data formats (NetCDF, HDF, GRiddedBinay), systems, platforms and technologies [37]. Besides, these data sources lack exposing their relat-ed data with enriched-semantics. The need for homogeneously accessing data sources aims at facilitating the integra-tion process which consists in the automated access, exploitation, and reuse of data for disaster prediction process.

2) EO APIs and RESTful-based services for interoperability: Some of the EO and meteorological systems provide their own related APIs for accessing, retrieving and managing data, others do not. Once APIs are unavailable, accessing data is not easily performed, as the user needs to download and store available data (i.e: FTP files), temporarily in a dedicated storage system. For these aforementioned issues, we chose to develop a data access service layer based on services to hide the heterogeneity of data access techniques. Web services govern the logical separation of concerns related to data, code and communication and enable users to perform remote calls over the Web and manipulate data through dedicated operations, through Application Programming Interfaces (APIs). The Open Geospatial Consortium has defined the OGC GeoAPI Standard [31] [32]. In this API, geospatial data are searched in a catalog of services via interfaces, bindings and applications. Another type of data services manipulating plain-text have been proposed in [26] [29]. This kind of API may cause latency issues when executing services with an increased volume of exchanged data. Besides, these services are unable to communicate with other EO APIs and systems. Furthermore, the Open Archives Initiative group has created the OpenAPI Initiative project (OAI) [35], which consists in documenting APIs using open meta-languages for APIs. This initiative relies on the specification of REST APIs (REpresentational State Transfer) [20] creating an open description format for API services vendor neutral, portable and open to be connected with other APIs. Our motivation to use REST is to propose a unified access to the existent APIs and to focus on data-driven based approach where we will be able to enhance data through semantics. Furthermore, semantics facilitate the linking mechanism between re-sources, make descriptions easily understandable allowing easier searching capabilities. To reach that, recent studies applied ontology-

[3] http://eopen-project.eu/

based approaches. In the next section, we highlight the related work to prediction systems and their related ontology-based approaches

B. Related work on ontology-based prediction systems

Many previous works have discussed why ontology is needed to solve decision making problems in disaster prediction systems. Devaraju et al. [7] presented an ontology-based approach to infer geographical events from sensor observations, by exploiting the ontological vocabularies with reasoning and querying mechanisms. However, this approach does not handle the data integration issue especially the volume and the variety of data. Llaves et al. [16] investigates how to infer and represent events from time series of in situ sensor observations. However, it does not support either additional sensor data sources nor additional data formats. Zhong [38] proposed a geo-ontology-based approach to decision-making in emergency management of meteorological disasters. It does not include other disaster types (like hydrological and geophysical disasters). The previously presented prediction approaches and several others face many common issues and limitations over earth observations big data. Most importantly, we noticed the impact of the huge volume and fast growth of EO data. Indeed, in the environmental monitoring domain, a large amount of information about climatological, meteorological, natural disasters, environmental processes are generated, which necessitate integration work. Also, the heterogeneity of EO data formats is a problem. In fact, data can be represented in different types (unstructured such as raster images and structured data such as databases). Finally, EO data integration and linking remain absent, while a good understanding of environmental phenomena needs collection and correlation from multiple data sources. For instance, making decision about floods and understanding this phenomenon requires integrating and analyzing infrastructure data, hydrological data, climatological data, etc.

C. Motivations

Access to EO data needs to be frequent and adequate means of monitoring EO data vigilance must be established to improve disaster prediction and management. However, multiple terms and formats are used describing data. Moreover, with the exponential amount of EO data, it is practically impossible to interpret data sources contents. Hereafter, we illustrate this statement with an example of two data sources that deal with the same concept (i.e, precipitation) but with heterogeneous data formats and access techniques. The first one is the Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS) which is proposed by the international organization (OSS). This data source provides data about precipitation, either daily, monthly or dekadly. Moreover, the presented data are formatted in BIL [6] and data are accessible through downloadable FTP links [8]. The second data source is the Current Weather Data [27], enacting the OpenWeatherMap API, proposed by the Open-Weather Company. The user accesses the API through its URL and obtains JSON-encoded weather information (including precipitation) being given geographic coordinates [27].

In this section, we noticed that for both data sources, access data techniques are different. For the first one, it would be mandatory for a user to temporarily store the FTP file, unzip it and try to understand the data, which are tedious tasks. And for the second one, data are accessed through the API URL. Therefore, it is necessary to unify access techniques and manipulate homogeneous formats. Moreover, it would be interesting if meaningful information is returned from both of data sources. More precisely, both data sources lack machine-readable semantic descriptions related to precipitation, such as the "observation measurements ", more information about the "location", the "clouds", etc. Otherwise, the user would be interested in obtaining additional information about the original source of data, or information related to the air matter composition of clouds.

Besides, our motivation consists in proposing a platform to seamlessly access and monitor heterogeneous data sources and integrate data with semantic annotations, in order to enhance interoperability between data sources and provide decision support for improving disasters prediction. The application of the PREDICAT process on the proposed example, in order to highlight the motivation and the advantages of the overall architecture design, will be more detailed in section 4.

III. PREDICAT ARCHITECTURE

Fig.1 presents the global architecture of the PREDICAT platform and its tiers. The layered architecture is composed of eight tiers, namely: (1) data collection layer, (2) big data layer, (3) data access service layer, (4) data processing layer, (5) semantic layer, (6) data integration layer, (7) application layer and (8) user interface layer.

A. Data collection layer

This layer encompasses different Web data sources relevant to earth observations and deals with different data format types (i.e: BILs, HDFs, NetCDF, etc.). For instance, CHIRPS uses satellite imagery to create gridded rainfall time series and data are downloadable via FTP links [6]. While the OpenWeatherMap, thanks to its current weather data API, provides the accurate weather for a given location. Other APIs are provided by the OpenWeatherMap source rendering 5day/3 hour forecast, 16 day/daily forecast, etc. NASA AIRS [30] (Atmospheric InfraRed Sounder) is collecting atmospheric, surface pressure and air quality data. The storage is essentially done on CMR (Common Metadata Repository) and the user can search data through the API CMR OpenSearch. NASA AIRS data are downloadable via a web link [19], where files are compressed in HDF formats. The Harmonized World Soil Database (HWSD) [36] is a raster database combining existing regional and national soils information worldwide. The HWSD database is downloadable through a dedicated link [11] where the rasters soil are formatted in .bil. Most of the data sources detailed in this layer are pointing to the heterogeneity related to their software applications and the used storage systems.

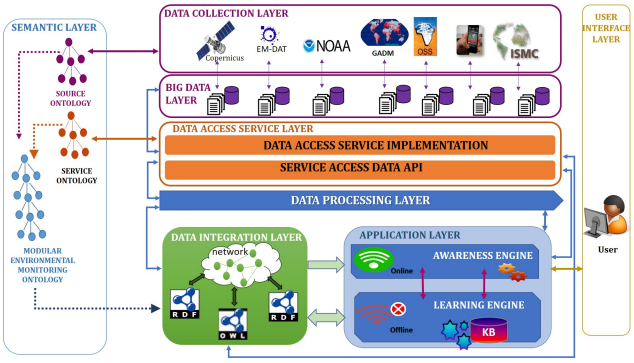

Figure 1. PREDICAT's layered architecture.

Moreover, data collection speed is variant in real-time, impacting on the update frequencies and the very rapid data-processing. The aim of this layer is to identify each source, describing its name, its features and its contents. Thus, facilitating the access when data are fetched form big data layer. Nevertheless, mining this large amount of data may consider a homogenized access without storing or replicating data temporarily. For all these encountered issues, we will present in the following sections the proposed solutions.

B. *Big data layer*

Since heterogeneous data sources generate large data types at unprecedented rate, they are stored in different dedicated storage systems as discussed in the previous layer. In this layer EO data are considered as raw data stored in different kind of structures related to the storage system of its source. Moreover in the big data scope [9], access to these data sources becomes a difficult task. Therefore, there is an urgent need for mechanisms providing a granular access to massive datasets. These large number of datasets in the big data field, may be subject to uncertainty, where data attributes are constructed using computational methods for collection mechanisms which are also subject to failure. These computational and statistical methods are covering the forecasting mechanisms or collecting sensor data. Therefore, data representations could vary from a data source to the other, causing untrustworthiness and contradictory meaning of data. Besides, the challenges of this layer relate to the access and the extraction of accurate EO data among the numerous datasets, in real-time, knowing the high update frequency changes. In addition, next layers try to unify data representations, where EO data will be rapidly and easily accessed through their dedicated storage systems. Otherwise, EO data will be fetched and accessed through services detailed in the next layer namely service layer.

Another challenging aspect for the PREDICAT architecture is how to interpret these datasets in order to get insightful information for environmental disasters' prediction. Next sections provide details on the management of these datasets.

C. *Service layer*

This layer encompasses two sub-layers:

1) Data acces service implementation: This layer deals with the implementation of services using the RESTful architectural style [4], [17], [20] and empowering these services with semantic descriptions about the services content and their related data. Thus, enabling to bridge the different knowledge representations across the heterogeneous multi-sources, detecting cross-processes relations and allowing reusability of services [24] across web-applications. In order to implement RESTful data access services, we have respected the set of best practices proposed in [3] such as, resources identification, resources manipulation using representations, and the usage of the Hypermedia HATEOAS mechanism. Following these recommendations, a set of RESTful services are developed using NetBeans as an Integrated Development Environment installed on a 64 bits machine. These services are then enriched with semantic annotations with linked data making them machine-readable and promoting the interoperability between heterogeneous resources. In order to set up these semantic annotations with linked data, our choice was set on Hydra Core Vocabulary [12], since it is a powerful vocabulary enabling the creation and management of Hypermedia-Driven APIs. The access to these APIs is detailed in the next sub-section.

2) Service access data API: The implemented services are published in a dedicated registry that defines standard API to ease the search of the annotated services description. Incoming requests are originated from PREDICAT users (i.e., OSS engineers, earth observation engineers or a simple user/citizen), which express their desired concepts (such as, temperature, precipitation, etc.), that should be matched with concepts in the registry. The semantic matching is based on user keywords provided as inputs to search on the registry. If there is a matching, the service URL or the list of sorted services is returned. This is achieved based on a matched score [22]. The next section focuses on semantic schemas related to the EO data, accessed by services access.

D. *Semantic layer*

This layer consists of 3 components, i.e., Modular Environmental Monitoring Ontology (MEMOn), source and service ontologies.

1) The Modular Environmental Monitoring Ontology (MEMOn): Several works proposed ontologies for the environmental monitoring domain, discussed in [18]. To address their problems, we propose MEMOn, a modular ontology for the environmental monitoring field, based on the upper level ontology Basic Formal Ontology (BFO) [2] and other existing ontologies such as the Common Core Ontologies (CCO), the Semantic Sensor Netowork ontology (SSN) and the ENVironment Ontology (ENVO).

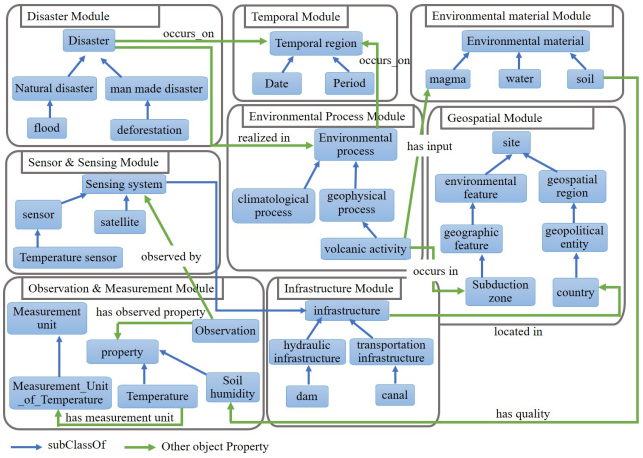

Figure 2. View of the MEMOn modules.

The proposed ontology consists of eight main modules covering the subdomains of environmental monitoring field, namely observation and measurement module, sensor and sensing module, disaster module, environmental process module, environmental material module, infrastructure module, temporal module and geospatial module, as illustrated in Fig.2. These modules incorporate all the different kinds of information entities handling all types of emergency situation, e.g., flooding, earthquakes, etc., spatial and temporal information, and sensing and observation information that are of importance to the environmental monitoring domain. The links between modules cover the relationships between sensing entities, observation entities and environmental events that they may cause, and also relationships arising from the fact that infrastructure objects like bridges, tunnel and dams could participate in environmental hazard factors. The main objectives of MEMOn ontology are 1) to share information in an environmental domain in a common vocabulary, 2) to ensure the semantic interoperability between heterogeneous sources and 3) to support the integration and linking of data together in order to build a global interactive network that permits to better understand environmental dynamics and natural phenomena. MEMOn was evaluated using quality metrics based on a set of criteria such as completeness, clarity, interoperability, etc. This evaluation was a worthwhile task since it permits to consider MEMOn as a high-quality ontology.

2) Source ontology: The vision of PREDICAT is to combine data generated from multiple monitoring systems such as Copernicus, OSS and NOAA and data provided by citizens. For rea-sons of data quality control and certainty, it is necessary to keep track of the provenance of the data. The provenance information in the environmental monitoring field is important since this information permits to assess data reliability and can be useful in decision-making. Thus, we propose to develop an ontology to represent such additional information of data. The sources ontology will include information about the immediate source of data, some measure of its authenticity or credibility, its relation to other data sources either they shared common products or not and the products it provides. The remaining entities in the sources ontology are created to encode relationships among sources and products given that a product may be derived from one or more other sources (e.g., precipitation data of a given region might be extract-ed from both OSS and NOAA sources).

3) Service ontology: The main purpose of service ontology is to enable the semantic representational knowledge inherent to services and their related relationships. The proposed ontology is managed through a dedicated framework that features different modules and interfaces, among which we cite the reasoner module. The latter relies on semantic relationships among services to perform inferences. These inferences are driven by rules, generating service composition schema and retrieving a newly inferred knowledge on services. For instance, consider two RESTful services S1 for querying temperature and S2 for querying precipitation. If the temperature value is around 15° and the value for precipitation is higher than 70mm then these two services should be complemented by the RESTful service S3 for querying the wind speed. The service ontology is subject to further work extensions taking into account services quality metrics to enhance service composition definition. Besides, it should be improved to consider interoperability issue between data access services. In fact, most of the times services are not compatible with each other. This makes interoperability a major issue for a successful data access service composition. Moreover, in order to execute services access to extract data, performance is handled by the next section discussing the data processing layer.

E. Data processing layer

The data processing layer deals with the services execution schema related to the user inquired concepts. Moreover, this layer tackles the following problems such that; reducing time-consuming processes, time-responses to fasten predictions and reducing costly-consuming bandwidth for requests. In fact, this layer uses the reasoner module and adds a new one named a decision maker. The latter takes as input the matched list of services for each concept and presents as an output the chosen service or combines it with others to compose a new service, based on its highest score of matching. It performs this task for each inquired concept mentioned in the user query. Besides, based on the service ontology as well as the selected services the reasoner proposes an orchestration schema for services accessing data. The main usage of the reasoner and the decision maker in the PREDICAT architecture is to select the viable service corresponding to the user request, thus, reducing the consumed response time and satisfying the user needs.

F. Data integration layer

The objective of data integration is to combine a large amount of data coming from various heterogeneous sources into a single consistent and global view of the data. One of the main problems of the data integration is the data heterogeneity. This variety can come from the structure and/or the format of the data and the vocabulary used to index the data. In general, each source has its specific characteristics. Several approaches have been proposed to cope with large-scale heterogeneous data integration [1], [15], [23]. Although the benefits of these approaches are obvious, modeling, linking and integrating EO data in such a way that capture the different representations of spatial and temporal contexts of observations that could be more explicitly modeled to improve data analysis still remains as a question.

As an alternative to these approaches to cope with this problematic, we propose to integrate all data sources to a common and global view with the target of augmenting the interconnections among data. The objective of this layer is not to copy and/or store data. In contrast, our aim is to virtually integrate and link heterogeneous data via the integration of their metadata. The proposed approach performs three steps which are extracting relevant entities from metadata, applying a virtual semantic annotation on the data and expanding the spatial and temporal contexts by relationships containing in MEMOn that supports flexible contextual spatial construction in terms of places, relationships between different representations of place and contextual temporal construction in terms of different representations related to a temporal setting. Finally, it stores the extracted and enriched information in a global RDF format. This latter offers a form of integration and query for the following layer.

G. Application layer

The application layer consists in two components, i.e. learning component and pre-diction component. The goal of the first one is to execute predictive models that learn from existing data to predict future trends, outcomes and behaviors. It takes the global RDF store as input and then generates new relations that helps to deduce knowledge and improve the performance of awareness. The prediction component handles the real time data and takes into considerations the inferred knowledge from the previous component to provide early warnings and decision support. Prediction systems have improved in recent years but they are not perfect yet. Early detection of natural disasters is still a need. In this layer, standard and rule-based reasoning would be employed (OWL reasoners and SWRL rule engine [10]).

H. User interface layer

PREDICAT users such that EO engineers or even ordinary users may have the possibility to query earth observations through the user interface layer. This layer is a front-end interface allowing to dialog with the PREDICAT architecture. In fact, the queried data sources will be displaying their related resulted data to end-users through this interface.

IV. Exemplar Use Case

The goal of this section is to provide an exemplar use case (see Fig 3) to illustrate how PREDICAT platform would accommodate a big data integration and give a decision-support for disaster predictions. Through our platform, two scenarios can be drawn. The first one illustrates the process after the detection/arrival of new EO data. The PREDICAT platform starts with implementing RESTful services to access the multi-source data from the Big data layer, if they are inexistent in the service registry. The RESTful services implementation is realized for each requested observation, in a way that a RESTful service can access and has the capacity to retrieve data from all sources generating this type of observation. Then, the semantic layer enhances the services with semantic enhancements thanks to the ontology repository (MEMOn, source ontology and service ontology), discussed previously. For example, the source ontology will link the observation "precipitation" with its originated data source OSS(CHIRPS) and its features, and the observation "temperature" with its originated data source the OpenWeatherMap and its features. Both sources and their characteristics are identified in the data collection layer and are mapped into the source ontology. Moreover, RESTful services are annotated with the Hydra vocabulary, thus generating Hydra annotated services descriptions. Afterwards, obtained data is fed into the data processing layer, which is responsible of the orchestration of services, when a service may compose with other services. Then, the outputted data is fed into the data integration layer which performs three steps. First, it extracts relevant information (such as temporal and spatial information) from metadata. Second, it links the observed properties extracted from the data such as precipitation and temperature with each other on the basis of object properties contained in the ontology repository and expands the spatial and temporal contexts by relationships defined in MEMOn. For instance, Paris is linked with its geographic coordinates in such a way that the system extracts all data in Paris with all different spatial representations. Finally, it stores the extracted and enriched information in a global RDF format. The obtained RDF store will contain all the necessary metadata and relationships required by the learning component in the application layer, in order to generate implicit knowledge. Indeed, based on the knowledge already contained in the learning component and the new data extracted from data sources, the learning engine explicits implicit knowledge and infer new knowledge which will be used later by the prediction engine.

In the second scenario, the user expresses in a query the desired observations (such as the temperature and the precipitation in Paris in 2018), through the PREDICAT user interface layer. The PREDICAT platform starts with searching and invoking RESTful services to access the multi-source data from the Big data layer. If necessary, it may invoke other services through the process of service composition.

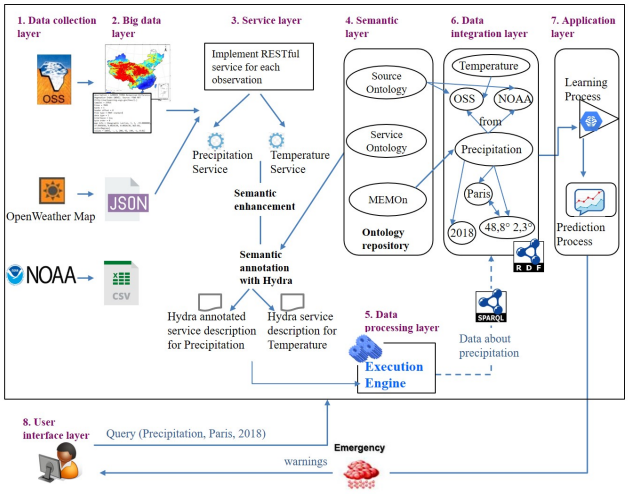

Figure 3. *An exemplar use case of the PREDICAT platform.*

At the same time, the user query is transmitted to the data integration layer in the form of a SPARQL query to interrogate the global RDF store and extract all knowledge and relationships related to the observations contained in the user query. From the result of this step and according to the rules contained in the learning layer, the platform executes a prediction process. For example, let us suppose that in the data integration layer, a heavy rainfall phenomenon which has occurred in 2018 in Paris is in relation with precipitation and temperature data observed in the same location and the same period of time. Then, on the basis of knowledge inferred from the learning engine the prediction engine predicts floods. As a result of the query submitted by the user: precipitation and temperature data in Paris, the PREDICAT platform provides all the data related to Paris, even if the spatial context representations in the data from OSS (CHIRPS) and OpenWeatherMap are different (country name or geographical coordinates) and an implicit knowledge which is flood warning.

V. CONCLUSION

In this paper, we proposed a semantic service-oriented platform to PREDIct natural CATastrophes. The PREDICAT platform aims at integrating and processing a large-scale heterogeneous big data generated from multiple sources, including that provided by citizens in order to provide decision support to effectively prevent against natural disasters. The contributions of our approach deal with 1) the use of ontologies to support semantic interoperability 2) the implementation of services that facilitate data access and extraction 3) the proposal of a data integration layer that ensures a global vision of EO data and its related spatio-temporal contextual information and 4) the proposal of a decision support system which allow predicting natural disasters. This work is still in progress. As future work, we intend to deal with data coming from social media such as twitter and real time messages sent by first responders or people in danger (including images and video). However, to use this data, it should be first analyzed by the respective analysis software. This step should also be considered. Then, we plan to deal with quality metrics related to services and data integration. On the one hand, services will be hosted and managed by a service provider consumed by EO customers. Moreover, quality measurements will improve the selected services for a better disaster prediction. On the other hand, computing the precision, accuracy, scalability and other measurements as data integration evaluation metrics will be necessary to demonstrate the effectiveness of our approach.

ACKNOWLEDGMENT

This work was financially supported by the "PHC Utique" program of the French Ministry of Foreign Affairs and Ministry of higher education and research and the Tunisian Ministry of higher education and scientific research in the CMCU project number 17G1122.

REFERENCES

[1] Abbes, H., &Gargouri, F.: Big data integration: A MongoDB database and modular ontologies based approach. Procedia Computer Science, 96, 446-455 (2016).

[2] Arp, R., Smith, B., Spear, A. D.: Building ontologies with basic formal ontology. MitPress.(2015)

[3] Berners-Lee, T., Hendler, J., Lassila, O.: The semantic web. Scientific American (Sci.Am), 284 (5), (May 2001), pp. 34-43

[4] Bülthoff, F., Maleshkova, M.: RESTful or RESTless – Current State of Today's Top Web APIs. The Semantic Web: ESWC 2014 Satellite Events, pp, 64–74 (2014).

[5] Copernicus program homepage,http://www.copernicus.eu/, last accessed 2018/06/29

[6] DATA CHIRPS Homepage, http://chg.geog.ucsb.edu/data/chirps/, last accessed 2018/07/17

[7] Devaraju, A., Kuhn, W., Renschler, C.:A formal model to infer geographic events from sensor observations. International Journal of Geographical Information Science, 1-27 (2015).

[8] FTP CHIRPS Products, ftp://ftp.chg.ucsb.edu/pub/org/chg/products/CHIRPS-2.0/, last accessed 2018/07/17

[9] Geppa, A., Linnenlueckeb, M.K., O'Neilla, T.J., Smith, T.: Big data techniques in auditing research and practice: Current trends and future opportunities. Journal of Accounting Literature (2018)

[10] Horrocks, I., et al.: Swrl: A semantic web rule language combining owl and ruleml. W3C Member submission, (21), 79. (2004).

[11] http://webarchive.iiasa.ac.at/Research/LUC/External-World-soil-database/HTML/

[12] Hydra Core Vocabulary Homepage, https://www.hydra-cg.com/spec/latest/core/#hydra-at-a-glance, last accessed 2018/07/17

[13] Kadadi, A., Agrawal, R., Nyamful, C., Atiq, R.: Challenges of data integration and interoperability in big data.*IEEE International Conference on Big Data (Big Data),* Washington, DC, pp. 38-40 (*2014).*

[14] Kaisler, S., Armour, F., Espinosa, J. A., Money, W.,: Big data: issues and challenges moving forward. In: 6th Hawaii international conference on system sciences (HICSS), 995–1004 (2013).

[15] Knoblock, C. A., Szekely, P.: Exploiting Semantics for Big Data Integration. AI Magazine, 36(1) (2015).

[16] Llaves, A., Kuhn, W.: An event abstraction layer for the integration of geosensor data. International Journal of Geographical Information Science (2014).

[17] Luo, Y., Puyang, T., Sun, X., Shen, Q., Yang, Y., Ruan, A., Wu, Z.: RestSep: Towards a Test-Oriented Privilege Partitioning Approach

for RESTful APIs. IEEE 24th International Conference on Web Services, 548-555 (2017)

[18] Masmoudi, M., Ben Abdallah ben Lamine, S., BaazaouiZghal, H., Karray M.H., Archimede, B.: An ontology-based monitoring system for multi-source environmental observations. 22nd International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (2018) (To appear).

[19] NASA EarthData Near Real Time Data, https://earthdata.nasa.gov/earth-observation-data/near-real-time/download-nrt-data/airs-nrt, last accessed 2018/07/17

[20] Neumann, A., Laranjeiro, N., Bernardino, J.: An Analysis of Public REST Web Service APIs. IEEE Transactions on Services Computing (2018).

[21] NOAA homepage, http://www.noaa.gov/

[22] Purohit, L., Kumar, S.: Web Service Selection Using Semantic Matching. Proceedings of the International Conference on Advances in Information Communication Technology & Computing (2016)

[23] Salmen, D., Malyuta, T., Hansen, A., Cronen, S., Smith, B.: Integration of intelligence data through semantic enhancement (2011).

[24] Selvakumar, G., kaviya, B.J.: A Survey on RESTful web services composition. International Conference on computer Communication and Informatics (ICCCI), (2016), pp, 1-4.

[25] Siddiqa, A., Karim, A., Gani, A.: Big data storage technologies: a survey. Frontiers of Information Technology & Electronic Engineering, 18(8), 1040-1070(2017).

[26] The Critical Zone Observatories Homepage, http://czo.colorado.edu/html/research.shtml, last accessed 2018/07/17

[27] The Current OpenWeatherMap API Homepage, https://openweathermap.org/current, last accessed 2018/07/17

[28] The Geoapi Homepage, http://www.geoapi.org/, last accessed 2018/07/17

[29] The Geoinformatics for Geochemistry System Homepage, http://www.earthchem.org, last accessed 2018/07/17

[30] The NASA APIs Homepage, https://earthdata.nasa.gov/api, last accessed 2018/07/17

[31] The OGC Public Engineering Report, https://portal.opengeospatial.org/files/?artifact_id=61224, last accessed 2018/07/17

[32] The Opengeospatial Homepage, http://www.opengeospatial.org/standards/netcdf, last accessed 2018/07/17

[33] The Opengeospatial White Paper, http://docs.opengeospatial.org/wp/16-019r4/16-019r4.html, last accessed 2018/07/17

[34] The Sahara and Sahel Observatory homepage,http://www.oss-online.org/en

[35] The W3C Recommendation Homepage for Data on the Web, https://www.w3.org/TR/dwbp/#documentYourAPI, last accessed 2018/07/17

[36] The World Soil Database homepage, http://webarchive.iiasa.ac.at/Research/LUC/External-World-soil-database/HTML/HWSD_Data.html?sb=4, last accessed 2018/07/17

[37] Vitolo, C., Elkhatib, Y., Reusser, D., J.A. Macleod, C., Buytaert, W.:Web technologies for environmental Big Data, Environmental Modelling and Software 63, elsevier (2015), 185-198

Zhong, S., Fang, Z., Zhu, M., & Huang, Q.: A geo-ontology-based approach to decision-making in emergency management of meteorological disasters. Natural Hazards,*89*(2), 531-554 (2017).