The ICARUS Ontology: A general aviation ontology developed using a multi-layer approach

Dimosthenis Stefanidis
dstefa02@cs.ucy.ac.cy
Computer Science Department, University of Cyprus
Nicosia, Cyprus

Chrysovalantis Christodoulou
cchris47@cs.ucy.ac.cy
Computer Science Department, University of Cyprus
Nicosia, Cyprus

Moysis Symeonidis
symeonidis.moysis@cs.ucy.ac.cy
Computer Science Department, University of Cyprus
Nicosia, Cyprus

George Pallis
gpallis@cs.ucy.ac.cy
Computer Science Department, University of Cyprus
Nicosia, Cyprus

Marios Dikaiakos
mdd@cs.ucy.ac.cy
Computer Science Department, University of Cyprus
Nicosia, Cyprus

Loukas Pouis
loukas.pouis@kpmg.com.cy
Management Consulting, KPMG Cyprus
Nicosia, Cyprus

Kalia Orphanou
kalia.orphanou@ouc.ac.cy
Cyprus Center of Algorithmic Transparency, Open University of Cyprus
Nicosia, Cyprus

Fenareti Lampathaki
fenareti@suite5.eu
Suite5 Data Intelligence Solutions Limited
Limassol, Cyprus

Dimitrios Alexandrou
dalexandrou@ubitech.eu
UBITECH Ltd.
Athens, Greece

ABSTRACT

The management of aviation data is a great challenge in the aviation industry, as they are complex and can be derived from heterogeneous data sources. To handle this challenge, ontologies can be applied to facilitate the modelling of the data across multiple data sources. This paper presents an aviation domain ontology, the ICARUS ontology, which aims at facilitating the semantic description and integration of information resources that represent the various assets of the ICARUS platform and their use. To present the functionality and usability of the proposed ontology, we present the results of querying the ontology using SPARQL queries through three use case scenarios. As shown from the evaluation, the ICARUS ontology enables the integration and reasoning over multiple sources of heterogeneous aviation-related data, the semantic description of metadata produced by ICARUS, and their storage in a knowledge-base which is dynamically updated and provides access to its contents via SPARQL queries.

CCS CONCEPTS

• **Information systems → Web Ontology Language (OWL)**; • **Computing methodologies → Knowledge representation and reasoning**; • **Applied computing → Aerospace**.

WIMS 2020, June 30-July 3, 2020, Biarritz, France

ACM ISBN 978-1-4503-7542-9/20/06…$15.00
https://doi.org/10.1145/3405962.3405983

KEYWORDS

ontology, aviation, datasets, services, queries

ACM Reference Format:
Dimosthenis Stefanidis, Chrysovalantis Christodoulou, Moysis Symeonidis, George Pallis, Marios Dikaiakos, Loukas Pouis, Kalia Orphanou, Fenareti Lampathaki, and Dimitrios Alexandrou. 2020. The ICARUS Ontology: A general aviation ontology developed using a multi-layer approach. In *The 10th International Conference on Web Intelligence, Mining and Semantics (WIMS 2020), June 30-July 3, 2020, Biarritz, France.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3405962.3405983

1 INTRODUCTION

The ongoing digitalization of every aspect of human activity has led to an exponential increase of digital data produced on a continuous basis from numerous sources. These data, broadly described as "Big Data," capture and represent activities, processes, and interactions between organizations and individuals, and create numerous opportunities for discovering unknown patterns, optimizing processes and products, devising innovative solutions, and improving decision making. These opportunities can be pursued thanks to a rich ecosystem of tools that emerged in recent years, comprising software, hardware, and algorithms that cover the whole spectrum from data production, extraction and communication to storage, processing and analysis. Taking advantage of the Big Data ecosystem offerings is a collaborative endeavor that requires the combination of knowledge, competences and resources from different sectors, along with specialised domain expertise.

However, the data sources lack standardization and incorporate varying data formats. As a result, data integration is a significant bottleneck to productivity. Furthermore, data integration and linking is a rather costly and underrated procedure, especially for SMEs, that do not have the necessary expertise and cannot invest the necessary time and resources to curate, interrelate, understand and share the primary and derived data with other interested parties, in a trusted manner. Typically data models are developed to manage the generated data by encoding the structure, format, constraints and their relationships to real-world entities. Unfortunately, aviation data providers employ quite different data models and thus aviation data models can vary along various dimensions [17]:

- *Data encoding format*: Data providers can use different formats to encode aviation data e.g. an airline carrier ID field could be stored as a three (IATA) or four letter code (ICAO).
- *Data field naming*: Field names assigned to values can be misleading and can lead to confusion without standardization e.g. a provider may use the name "AT" while another may use "arrTime" for the field "aircraft arrival time".
- *Data semantics*: Even if two data fields are identical, that doesn't ensure that the data represents the same information e.g. the meaning of "aircraft arrival time" may be different across providers as it may correspond to a scheduled or an actual arrival time.
- *Spatial and temporal resolution*: Data values may be recorded at different temporal frequencies (e.g. once per hour) or spatial regions (e.g. airspace sectors, geographic regions).
- *Measurement unit conventions*: Units are often omitted in the data storage schemes and can lead to problems when different units are employed across different systems e.g. metric vs imperial, feet vs flight level.

To deal with the data integration challenges, many recent efforts focus on integrating Big Data systems and tools into domain-specific platforms that can attract resources and stakeholders, allowing the seamless combination of competences and data sets, to develop and offer added-value services and novel solutions. ICARUS [8] represents such an effort that focuses on developing a novel Big Data platform for the aviation domain. The ICARUS platform leverages a variety of proprietary and open datasets related to the aviation domain, and seeks to help stakeholders whose operations are directly or indirectly linked to aviation to enhance their data reach and share or trade datasets, intelligence, and expertise in order to gain better insights, improve operations, and increase passenger safety and satisfaction. ICARUS aspires to operate as a multi-sided platform that will allow exploration, curation, integration and deep analysis of original, synthesized and derivative data, characterized by different volume, velocity and variety, in a trusted and fair manner [8].

To meet this goal, a Big Data platform like ICARUS needs to provide its users with user-friendly services for discovering knowledge about the platform's assets: datasets, tools, algorithms, usage statistics, data quality patterns, registered experts etc. As these assets evolve dynamically, for instance with the frequent update of registered data sources or the replacement of older techniques with improved algorithms, the platform must be designed to support its own evolution: new datasets need to be semantically aligned with data already captured by the platform, the platform needs to integrate seamlessly changes in datasets served, data providers should have adequate tools to manage their assets on-demand, and users should be notified according to interests registered explicitly in the platform or inferred automatically by intelligent algorithms. To address these requirements, we explore the use of ontologies as a tool for formally describing various information resources integrated inside and managed by the ICARUS platform. Ontologies represent a powerful tool for defining and describing formally and explicitly a domain of discourse in terms of its concepts, their features, their relationships, and the various syntactic and logical constraints they abide to [3]. Ontologies are used for knowledge representation, data integration and decision making [27]. Ontology languages and ontology-based tools have been developed [23] [28] to engineer semantic information systems in various domains [33] [32] [6] [1] [29].

In this paper, we present our work in the study of existing aviation-related ontologies and the applicability thereof in developing semantic components for the ICARUS platform. We introduce the ICARUS ontology [1], which adopts, integrates and extends domain ontologies to represent semantically key concepts of the ICARUS information system. We show how we use the ICARUS ontology to expand the ICARUS platform with components that allow its users to define and execute semantic queries about the offered datasets and services, and to receive recommendations about offerings of interest. Finally, we describe a number of use-cases demonstrating the functionality and flexibility provided by proposed ICARUS ontology. The main contribution of the ICARUS ontology is to identify, reuse and extend existing domain ontologies with additional new concepts and relations in order to build a novel integrated domain aviation ontology.

The remaining of this paper is organised as follows. Section 2 provides an overview of the ICARUS platform. In Section 3, we describe the design process of the ICARUS ontology and its class hierarchy. In Section 3, the evaluation of the proposed ontology is performed using SPARQL queries in three application use case scenarios. The first use case scenario deals with the representation and modelling of aviation-related data extracted from Twitter and their sentiment score through the ICARUS ontology. The second use case scenario focuses on the representation and modelling of assets (data or service assets) related to epidemics (e.g. COVID-19) that can be integrated with aviation-related data through the ICARUS ontology in order to improve the forecasting capabilities of epidemics models. The third use case scenario deals with the use of a recommendation algorithm integrated with the ICARUS ontology for recommending datasets and services based on user's preferences. In Section 5, we outline the related work on applications of ontologies in the aviation domain. Finally, the conclusion is given in Section 6.

2 THE ICARUS PLATFORM

The design of the ICARUS platform followed a requirements analysis process, driven by a set of "user stories" that recorded usage scenarios emerging from a variety of application areas and user groups [9]. Based on this analysis, ICARUS is expected to provide services to four user categories: data providers, who contribute datasets to the platform; data consumers, who consume

[1]ICARUS ontology is freely available at: https://github.com/UCY-LINC-LAB/icarus-ontology

data retrieved from the platform through its services; service asset providers, who contribute machine learning, data analytics, visualization, and other software services on top of the ICARUS data-value chain; service-asset consumers, who take advantage of services deployed through ICARUS, and administrators, who manage and monitor the platform. As expected, many of the functional and technical requirements identified in the requirements analysis process, relate to services aiming at enriching ICARUS assets with semantic information that facilitates: data integration and categorization; exploration, querying and keyword-based search, and the provision of intelligent recommendations to ICARUS users regarding available data and services, the usage thereof and relevant best practices recorded.

To implement and deliver the identified functional and technical requirements, the ICARUS project developed a reference architecture comprising key software components and their interrelationships. This architecture is conceptually divided in three main tiers:

- The *On Premise Environment*, which runs on the data provider's site and allows data providers to prepare their datasets (especially private or confidential ones) in order to be uploaded in the ICARUS platform.
- The *Secure and Private Space*, which comprises dedicated virtual machines that are spawned on demand to execute user-defined analytics jobs in an isolated and secure environment. The Secure and Private Space comes with capabilities for encryption, decryption, the handling of keys etc [7].
- The *Core ICARUS platform*, which hosts the modules that offer the core services of ICARUS: the user front-end with its analytics and visualization workbench; resource orchestration; data management, storage, indexing, exploration and query; access control and policy management; application management, monitoring and usage analytics, and recommender system.

The ICARUS ontology aims at facilitating the semantic description and integration of information resources that represent the various assets of the ICARUS platform and their use. To this end, the ontology should enable the integration and reasoning over multiple sources of heterogeneous aviation-related data, the semantic description of metadata produced by ICARUS, and their storage in a knowledge-base which is dynamically updated and provides access to its contents via a query (see Figure 1). More specifically, the core uses of the ICARUS ontology are:

(1) To facilitate the semantic annotation of datasets in order to capture the structural and semantic characteristics of the various entities in each given dataset.
(2) To represent semantically other entities of the ICARUS platform such as service assets, deployed algorithms and their popularity, assets popularity and user's interactions.
(3) To drive the extraction of metadata from ICARUS platform operations, their semantic representation, and their storage in the ICARUS knowledge-base.
(4) To support the continuous integration of new datasets, services, and human experts into the platform, while maintaining the ICARUS data model and the ICARUS knowledge-base up-to-date.
(5) To support the search and query over multiple data sources and information assets available on ICARUS and on other open aviation-related datasets, such as open data, epidemics data and data extracted from Twitter or other online social networking sites.
(6) To provide an application-programming interface to feed the algorithms of the ICARUS recommendation engine with useful information.

The intended users of the ICARUS ontology will be: a) data providers and consumers who are the key stakeholders in the aviation value chain industry, b) IT industry players supporting the aviation value chain, such as IT companies, web entrepreneurs and software engineers, c) universities, research organizations, and policy makers who study the aviation ecosystem and its multifaceted impact, and d) the general public e.g. passengers.

3 THE ICARUS ONTOLOGY

The amount of available data in the aviation domain has been increasing over the last decades, since data providers have begun publishing their data through various digital platforms. In order to structure and analyze these data, an ontology is needed to organize the concepts and to define the interrelationships that exist for the aviation domain.

According to Grimm et al. [27], ontologies can be differentiated into the following types: (i) top level ontologies that consist of general and abstract concepts which can be imported into other ontologies (i.e. notions of time or space); (ii) domain or task ontologies which represent the knowledge about a specific domain (i.e. the aircraft) or a general task (i.e. cooking); (iii) application ontologies, which can represent and refine specific aspects of domain ontologies that can be used in a specific application considering specific usage scenarios.

An aviation domain ontology should be able to promote data-driven collaboration between the domains that are directly or indirectly linked with the aviation sector, bringing together stakeholders from diverse domains such as Aerospace, Retail and Weather. In this paper, we introduce a novel aviation domain ontology, called the ICARUS ontology, which is designed based on a multi-layer approach. The ICARUS ontology has been developed as a key component of the ICARUS system platform [2] for integrating and semantically enriching aviation-related data in different formats and from different data sources. The ICARUS ontology, however, can also be used as a standalone domain ontology. The ICARUS platform aims at building a novel data value chain in aviation-related sectors, driving data-driven innovation and collaboration across currently diversified and fragmented industry stakeholders, acting as multiplier of the 'combined' data value that can be accrued, shared and traded, and modernizing existing processes in the aviation domain. Using methods such as big data analytics, semantic data enrichment, and blockchain-powered data sharing, the ICARUS platform aspires to address critical barriers for the adoption of Big Data in the aviation industry (e.g. data fragmentation, data provenance, data licensing and ownership, data veracity). Through the ICARUS platform, aviation-related companies, organizations and researchers will be able to explore, curate, integrate and analyse

[2] https://www.icarus2020.aero

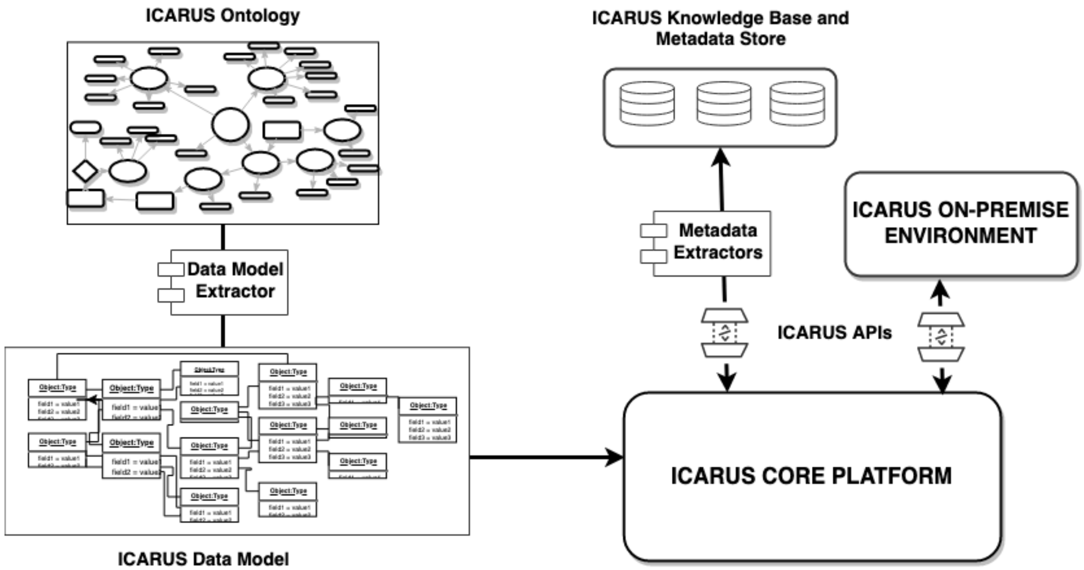

Figure 1: Ontology and Knowledge-Base Extension of the ICARUS Platform Architecture.

original, synthesized and derivative data characterized by different velocity, variety and volume in a trusted and fair manner.

The ICARUS ontology has been designed in order to enable the mapping of aviation-related datasets to the ontology concepts and also to answer general competency questions that can be useful to the aviation industry domain. For the development of the ontology, the Protégé software tool [23] and the OWL/XML syntax have been used.

3.1 Multi-layer Approach for the Ontology Development

To design a domain ontology, one can adopt various methods. The most known examples are either to extend existing ontologies or to develop the ontology from the ground up. In our work, we use a combination of these methods by using a multi-layer approach so that the ICARUS ontology can represent both metadata and aviation-specific concepts. This approach has firstly been proposed for the development of the CAMeOnto [2] ontology, a general meta-ontology for context modelling. As displayed in Figure 2, the top-level ontology (C ontology) describes the meta-contexts and attributes related to metadata of datasets and allows the addition of specific domain context ontologies. Other independent aviation domain ontologies (i.e. C_1, C_2, C_3, C_4) can be included under the top-level ontology (and even extend the ontology) to describe various (aviation-related) concepts. The entities and properties of each independent domain ontology should be integrated with the rest of the domain ontologies that are incorporated under the top-level ontology.

In our case, C_1, represents the NASA ontology [18] which has been selected since it includes most of the concepts related to aviation. The NASA ontology has been incorporated under the top-level ontology and it has been expanded with new concepts, data fields and relationships. The representative entities modeled within the

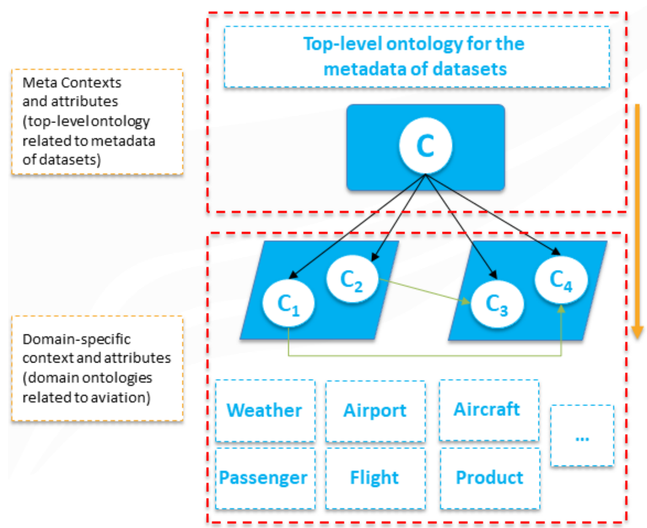

Figure 2: Multi-layer approach of the ICARUS ontology where C is the top-level ontology and C_1 - C_4 are independent aviation domain ontologies

NASA ontology are the following: *Flight* (e.g. flight plans and radar flight tracks, etc.), *Aircraft* and manufacturers (e.g. aircraft characteristics, models, etc.), *Airport* and infrastructure (e.g. runways, taxiways, terminals, gates, etc.), *Airline* (e.g. air carrier name, country of registry, etc.) and entities that represent air traffic management initiatives (TMIs) (e.g. ground delay programs, ground stops, reroutes, etc.). The new concepts that expand the NASA ontology have been incorporated under the ICARUS domain ontology (i.e. *Icarus extra properties*).

Another ontology that we used is the Epidemiology Ontology (EPO) [24], which describes epidemiologically relevant concepts such as transmission mode, epidemiological parameters and demographic parameters. It contains concepts that could facilitate the linking between epedimiological data and aviation-related data. By improving the level of detail and the linking capabilities between aviation-related data and health-related data is likely to lead to more accurate epidemic predictions. The EPO ontology has been incorporated under the top-level ontology (C3) and it has been expanded with new data fields and relationships that connect the aviaton-related entities with the concepts related to aviation (e.g. flights). Finally, the new concepts that expand the EPO ontology have been incorporated under the ICARUS domain ontology (i.e. *Icarus extra properties*).

The multi-layer approach has many benefits such as: a) Modularity and interoperability: The ontology is organized in a modular manner since the top-level ontology and each independent domain ontology are arranged in different modules, which facilitates the ontology's maintenance and extendibility and b) Flexibility/Scalability: The multi-layer approach ensures maximum ease of use for retrieving and querying data.

In addition to the multi-layer approach, another related ontology from a diverse domain is also integrated with the ICARUS ontology i.e. the DCAT ontology [21], which refers to the documentation (e.g. license) of the ICARUS Ontology and of each Dataset Entity/Concept. The DCAT ontology defines three main classes: the 'Catalog' which represents the index of the dataset; the 'Dataset' which represents a dataset in a catalog and the 'Distribution' which represents an accessible form of a dataset e.g. a downloadable file, an RSS feed or a web service that provides the data.

3.2 Definition of Classes and Class Hierarchy

The ICARUS ontology contains a total of 382 classes, 151 object properties and 450 data properties. The top-level ontology consists of 16 new classes, as displayed in Figure 3, while it consists of six main representative entities, which have been introduced to the ICARUS ontology:

- *Asset*: This class represents the generalized entity of a data or a service asset. It consists of properties that describe different types of assets e.g. license, publication date, title, description, categories, etc.
- *Data Column*: This class represents the column id of the columns on each provided data asset. Furthermore, it contains attributes that describe the quality of a data column e.g. percentage of missing values, percentage of duplicate values, etc.
- *Data Aviation-related type*: This class represents the generalized structure of defining the possible columns that can be provided in any aviation-related data asset e.g. *Aircraft*, *Airport*, *Flight*, *Bond* etc.
- *Data Value type*: This class would be useful for enabling the ICARUS ontology to semantic annotate data instances of diverse format types. It represents the general value types that the instances of the corresponding columns of an aviation-related data asset can take e.g. count value, free text values, geo coordinates value, geo name value, key value, measured value and time point value.
- *ML Algorithms*: This class represents the machine learning algorithms that are available in the ICARUS platform. Furthermore, it is connected with the service assets, as each service can contain one or more machine learning algorithm(s), and with the users in order to obtain usage statistics for each algorithm.
- *Platform User*: This class represents the users of the ICARUS platform. It consists of data properties like organization's name and preferences, and object properties that describe several relationships between users, assets and algorithms e.g. userViewedAsset, userPurchasedAsset, etc.

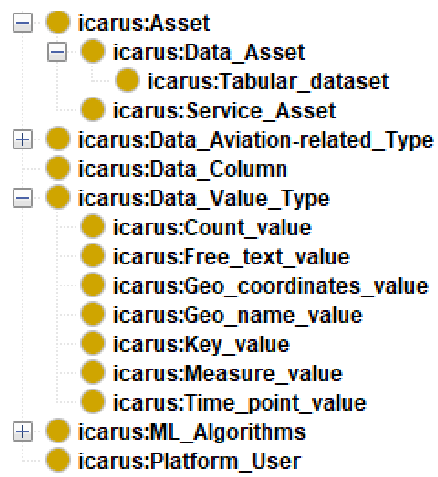

Figure 3: ICARUS Ontology Class Hierarchy

As previously mentioned, the top-level ontology represents attributes related to metadata of datasets. As displayed in Figure 4, a provided dataset (classes *Dataset* and *Tabular dataset*) consists of columns (class *Column*) where each column might take instances which are related to common aviation concepts (class *Aviation related type*), otherwise, it would be matched with a general value type (class *Data Value type* e.g. count value, free text value, time point value). The classes from the NASA ontology, are represented under the *NASA Ontology* while the "Icarus extra properties" represent the additional concepts that expand the NASA ontology. Both ontologies have been incorporated under the *Aviation related type* class, as displayed in Figure 5. Some of the additional classes are defined as equivalent to a class from the NASA ontology, since the specific class has already been defined in the ICARUS ontology e.g. the *AirCarrier* class is equivalent to the *AirCarrier* class from the NASA ontology.

At the current version of the ICARUS ontology, the *Icarus extra properties* includes the classes: *Baggage*, *BaggageBelt*, *Bond* and *BondLoading*, *Place*, *Country*, *City* and *State* where their data properties that have been added to this ontology have been connected with the NASA *Airport* class. Also, it includes the classes: *Product*, *Sales*, *InflightSales*, *Person*, *Passenger* where their data properties that have been added to this ontology, have been connected with

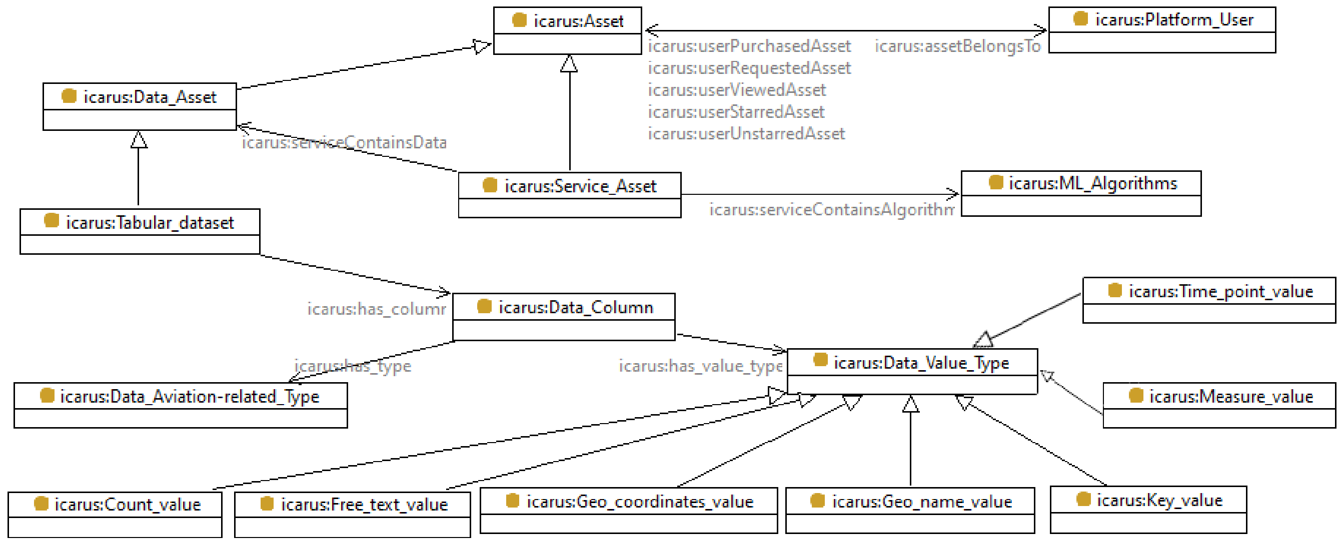

Figure 4: Relations of the main classes and sub-classes of the top ontology

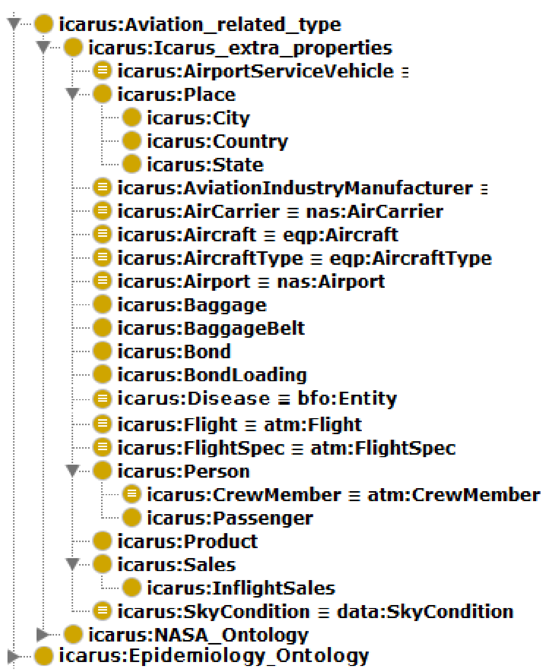

Figure 5: Data Aviation-related type subclasses

the NASA *Flight* class. In addition, some other classes have been added as equivalent to the corresponding classes from the NASA ontology such as: *AirportServiceVehicle*, *AirCarrier*, *Aircraft*, *AircraftType*, *CrewMember* and *SkyCondition*. An aircraft infrastructure component encompasses all aspects of the associated airport infrastructure i.e. terminals, gates, baggage belts, runaways, taxiways etc. Therefore, the *BaggageBelt* is represented as a subclass of the *AircraftInfrastructureComponent* that is incorporated from the NASA ontology. Each flight and airport are also associated with a baggage belt, thus, the *Flight* and *Airport* classes are connected to the *BaggageBelt* class using the relations 'associated with BaggageBelt' and 'has a BaggageBelt' respectively. The airport can represent either a departure, arrival or an associated airport, associated with the corresponding subclasses.

The bond loading in a flight includes the loading of diverse categories of products i.e. inflight sales such as food, snacks, drinks, and duty-free products, airport duty-free products, safety and entertainment equipment. Therefore, the class *BondLoading* is associated with the classes *Product* and *Bond* using the relations: 'has_product' and 'partOfBond' respectively. The class *Flight* is associated with the classes *Bond* and *InflightSales* using the relations 'belongstoBond' and 'hasInflightSales' respectively. The *InflightSales* is a subclass of *Sales* that is connected to the *Product* class with the relation 'hasProduct'. The class *Bond* also includes some general statistics regarding the products loading in the flights. Moreover, the *Person* class represents both the persons travelling on board including crew members and passengers (*CrewMember* and *Passenger* are subclasses of the *Person* class), but it can also include any other person who works on the airport. The demographics of the person such as: age, citizenship, date of birth, ethnicity, gender, marital status, nationality are assigned as data properties to the class. The class *Flight* is associated with the *Passenger* class with the relation 'hasPassenger' and with the *CrewMember* class with the relation 'hasCrewMember'. Each person may carry one or more baggages on the plane. Therefore, the *Passenger* class is connected with the *Baggage* class with the relation 'ownsBaggage'.

Also, every person who either works in the airport or on board or a passenger are assigned with their birth location (whenever its known). The location is represented with the *Place* class which has three subclasses: the *Country*, the *City* and the *State* where the city 'belongsTo' the country and the country is 'partOf' a state. The class *Person* is associated with the *Place* with the relations 'livein' and 'birthplace'. The *Place* class can also represent the location of the airport, thus it is also associated with the *Airport* and the *MeteorologicalReport* class with the relations 'AirportGeographicPlace' and 'ForecastingGeoPlace' respectively. The *Place* is also connected with the *AviationIndustryManufacturer* with the relation 'manufacturerPlace'.

3.3 ICARUS Ontology Extension

An exhaustive list of fields and relationships describing thoroughly all the datasets that can emerge as potentially relevant to the aviation domain cannot be drafted through the ICARUS ontology. Consequently, the ontology represents a dynamically changing representation of the evolving aviation data landscape that can easily be extended and updated.

To extend the ICARUS ontology, new individual domain ontologies could be added under the top-level ontology. For example, existing domain ontologies from diverse aviation-related domains such as transportation [10] [20], tourism [14] or health [25] can be added as independent domain ontologies (C_3, C_4, C_5 ...) and their entities/properties can be connected to some of the already incorporated domain ontologies (C_1 - C_2,..).

Moreover, a novel domain ontology with new concepts and relations of any particular domain, can also be integrated with the ICARUS ontology as an independent domain ontology. To achieve that, its new classes and properties should be associated with one or more of the already incorporated domain ontologies (C_1 - C_2,...). Examples of such ontologies are the Twitter and Emotion ontologies described in Section 4.2, for representing twitter data and performing the sentiment analysis.

In case that any of the recently-added domain ontologies include one or more classes that are already defined in the existing integrated domain ontologies, the two classes should be defined as equivalent [16]. For instance, if the AIRCRAFT ontology [4] would be incorporated into the ICARUS ontology, the *Aircraft* class from the AIRCRAFT ontology would be defined as equivalent to the *Aircraft* class from the NASA ontology.

4 EVALUATION AND USE CASES SCENARIOS

In this work, we use SPARQL [15] queries to retrieve information relevant to the concepts of the ontology and to answer a sample of the competencies questions. Furthermore, we store the data in a triplestore (also known as RDF store) called Virtuoso [12] [13], a high-performance and scalable Multi-Model RDBMS which outperforms other triplestore in terms of query-based performance [5] [30] [22]. Next, in order to evaluate the functionality of the ICARUS ontology, we describe three use case scenarios.

4.1 Reasoning using SPARQL Queries

To perform the reasoning in the ontology, we used three testbed that consist of datasets and/or services. The first testbed consists of the dataset "Delayed flights" (Dataset_1) which contains information regarding the delayed flights. In particular, the dataset contains the following columns: flight code, departure airport, arrival airport, departure terminal, departure gate flight, departure scheduled time and flight departure delay time in minutes. The second testbed consist of the dataset "Airline flight data" (Dataset_2) and represents the flights' data of different airlines. Specifically, the dataset contains the following columns: flight code, departure airport and arrival airport (as in Dataset_1) and number of total passengers in the flight. Figure 6 displays the column and row instances of the Dataset_2 and how these are mapped to the ontology concepts. Dataset_2 is a tabular dataset with four columns. Thus, Dataset_2 is an instance of the *Tabular_Dataset* class and the four columns: *flight_code*, *total_passengers*, *arr_airport_code* and *dep_airport_code* are instances of the *Column* class. The instances of the columns *flight_code* and *total_passengers* are mapped to the *Flight* class, while the row instances of the columns *arr_airport_code* and *dep_airport_code* are mapped to the *Airport* class. The two row instances of the dataset include details about two flights and their corresponding arrival and departure airport. The object properties that are introduced for the ontology mapping consists of: 'has_column', 'has_type', 'departureAirport' and 'arrivalAirport'. For instance, Dataset_2 'has_column' *flight_code* and the column *flight_code* 'has_type' *Row_Flight* which is an instance of the *Flight* class. The *Row_Flight* is connected with an arrival and destination airport which are instances of the class *Airport*. Finally, the third testbed consists of 3 datasets ("Coronavirus Cases", "Passengers Data", "Airline Flight Data") and 2 services ("Coronavirus Cases Prediction", "Predict Transmission of Coronavirus using Passengers' Data").

We perform the following SPARQL queries for answering a sample of the competencies questions:

- Question 1: ''Which datasets contain columns about flight delay time?'' To answer Question 1, we run the SPARQL query, as displayed in Listing 1, using as input all the three testbed datasets. The result of Question 1 is the dataset Dataset_1: Delayed flights.
- Question 2: ''Which is the airport departure terminal for a specific flight?'' To answer Question 2, we run the SPARQL query, as displayed in Listing 2, using as input only the first testbed dataset (Dataset_1). The result of Question 2 is the departure terminal number: 3.
- Question 3: ''How many were the occupied seats on a specific flight?'' To answer Question 3, we run the SPARQL query, as displayed in Listing 3, using as input only the second testbed dataset (Dataset_2). The result of Question 3 is the total number of 150 flight passengers.

```
SELECT
   DISTINCT (str(?id) AS ?dataset_id)
   (str(?name) AS ?dataset_name)
WHERE {
   ?dataset rdf:type icarus:Tabular_dataset .
   ?dataset icarus:has_column ?column .
   ?type rdf:type atm:Flight .
   ?type icarus:flightDepDelayMin ?value .
   ?column icarus:has_type ?type .
   ?dataset icarus:dataset_ID ?id .
   ?dataset icarus:dataset_name ?name .
}
```

Listing 1: SPARQL query for datasets that have a column for "flight delay time"

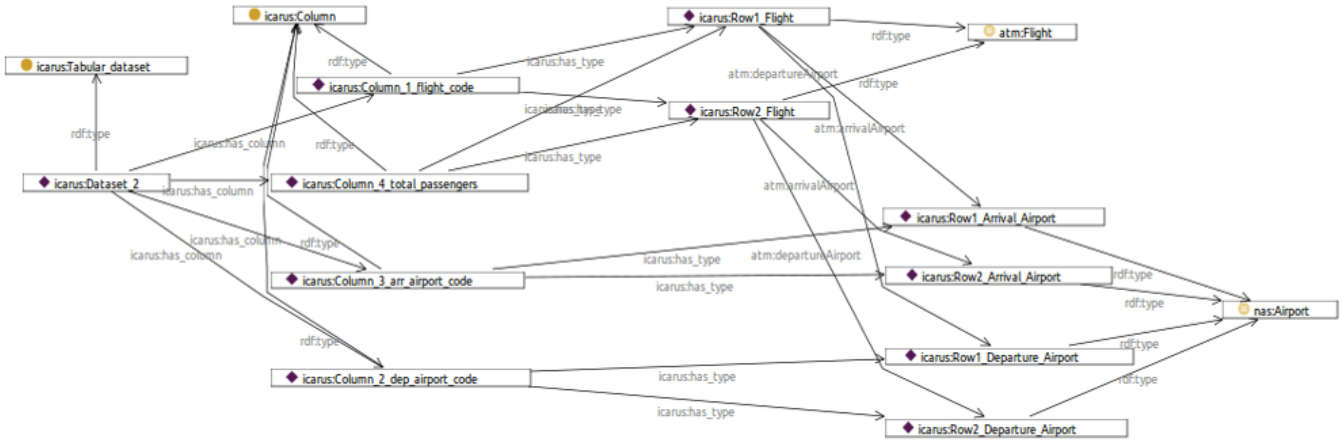

Figure 6: Dataset_2 mapping with the ICARUS ontology

```
SELECT
    (str(?terminal_number) AS ?departure_terminal_number)
WHERE {
    ?flight_code icarus:flightCode "DD121"^^xsd:string .
    ?flight_code icarus:departureTerminal ?terminal .
    ?terminal nas:terminalID ?terminal_number .
}
```

Listing 2: SPARQL query for finding the departure terminal of a flight

```
SELECT
    (str(?passengers) AS ?flight_total_passengers)
WHERE {
    ?flight_code icarus:flightCode "DD121"^^xsd:string .
    ?flight_code icarus:flightPassengers ?passengers .
}
```

Listing 3: SPARQL query for finding the occupied seats in a flight

4.2 Use Case Scenario: Application of the ICARUS Ontology to Twitter Data

The basic idea behind the proposed use case is to utilize the ICARUS ontology for providing more specific categories of sentiment scores for the aviation-related tweets. The aim is to use the ontology and a set of tweets related to the aviation domain to answer some competencies questions based on the sentiment scores for specific entities, e.g. airline. In order to achieve this, our approach would be completed in three main phases: (a) to retrieve tweets based on the concepts/entities stored in the ontology, (b) to extent the domain ontology with new concepts related to the retrieved tweets, and (c) to perform sentiment analysis on a set of retrieved tweets, by including emotion categories in the ontology. More specifically, the following steps have to be followed:

- Data Collection which includes the retrieval of tweets via the Twitter Streaming API using as search terms a combination of the entities in the ontology. In order to be compliant to GDPR requirements[3], we do not store any user-specific information. The data would be collected using three methods: a) by collecting tweets that are published in the geolocation coordinates of the airport, b) by collecting tweets based on hashtag & mentions of airport and airlines and c) by collecting tweets from airlines accounts. All this information i.e. airport geolocation, airline and airport names, airline/airport twitter user account will be retrieved from the ontology instances in the knowledge base. Therefore, the ICARUS ontology will be extended with a new ontology, the Twitter ontology, that will include the following concepts: the twitter user account, the number of followers and the tweet. The properties of these classes will be connected to the *Airline* and *Airport* class from C_1 ontology. Examples of retrieved tweets are displayed in Table 1.
- Data Pre-processing which includes the application of data cleaning and natural language processing (NLP) techniques to the retrieved tweets.
- Ontology Extension which includes the mapping of the concepts from the retrieved tweets with the existing ontology attributes and the addition of new subclasses and properties under the top-level ontology. In the current setup of the ICARUS ontology, classes that represent geo-location of the airport are already incorporated into the NASA ontology. The timestamp information, the tweet, the number of followers and the twitter user account will be added to the Twitter ontology [11].
- Sentiment analysis which has to be performed on each of the retrieved tweets. The sentiment scores would be assigned not to whole statements (i.e. tweets) but to the various relative concepts i.e. airlines, airports, flights. The sentiment scores can be computed independently of the ICARUS ontology and then the score would be stored in the Twitter ontology or

[3]European Union. 2016. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). Retrieved July 10, 2019 from *http://data.europa.eu/eli/reg/2016/679/oj*

the ICARUS ontology would be enriched in order to include all the possible emotion categories. In the second case, the corresponding emotions from the ontology of emotional categories will replace the sentiment score as in [11]. Examples of retrieved tweets and their estimated sentiment score is displayed in Figure 7.

A: Tweets collected using the location of the airport	I am at Los Angeles International Airport.
B: Tweets extracted using airport mention	Cleared: Construction on JFKennedyExpressw..
C: Tweets extracted from airline's twitter user account	Japan Airlines 787 rotating out of RWY 27R JP.

Table 1: Sample tweets used for the extension of the ontology.

full_text	id	sentiment1
I'm at Los Angeles International Airport - @fl...	5cd0faecead3c34fc153a345	0.0000
Ready to fly! See you on the other side! #hone...	5cd14bdbead3c34fc153a836	0.3612
Japan Airlines 787 rotating out of RWY 27R JP✈...	5cd197d9ead3c34fc153af07	-0.1677
Cleared: Construction on #JohnFKennedyExpressw...	5cd1be63ead3c34fc153b4d8	0.1027
Every time I walk past this, I giggle. Everyti...	5cd0f75dead3c34fc153a2ff	0.5023

Figure 7: Extracted aviation-related tweets with their sentiment score

Some indicative questions that the ontology would be queried when applying to Twitter data are:

- Which is the most popular airline? (searching for the airline with the most positive sentiment)
- Which airline has the lowest popularity? (searching for the airline with the most negative sentiment)
- Which airport is mostly preferred for travellers? (searching for the airport with the most positive sentiment)
- Which airport is the worst based on travellers' preferences? (searching for the airport with the most negative sentiment)
- Which is the most popular airline/destination airport on a specific month? (searching for the airline/airport with the most positive sentiment considering only the tweets of the particular month)

4.3 Use Case Scenario: Application of the ICARUS Ontology for Epidemics Data

Some of the current challenges of health organizations are to locate, collect, explore and integrate reliable data about airline and human mobility, with a sufficient geographical coverage and resolution. Improving such level of detail would result in more accurate epidemic predictions and a possible estimation of relative revenue losses to be expected in different pandemic scenarios by estimating the reduction of travellers on the airline mobility network (changes in air passenger in/outflow to/from specific regions or countries).

The ICARUS ontology and the relationships between each entity can be utilized in order to combine epidemics data with other aviation-related data for data analytics and epidemic forecasts. For instance, aviation-related data such as flights, travel restrictions, population, travellers' age, gender, income, can be integrated and utilized in order to find mobility and interactions patterns across different individuals and countries. Furthermore, the ICARUS ontology could be utilized for finding analytic tools and services that combine datasets and machine learning algorithms for predicting the spreading of the diseases.

In this use case, the objective is to utilize the ICARUS ontology and a set of datesets related to the health and aviation domain to answer some competencies questions for several entities, e.g. airline, country. In order to achieve this, our approach would be completed in three main phases: (a) use the ICARUS ontology to retrieve aviation-related data and health-related data based on the concepts/entities stored in the ontology, (b) integrate the datasets based on their relationships as defined in the ICARUS ontology, and (c) use SPARQL queries to extract new knowledge and insights from the combined dataset.

Some indicative questions that the ontology would be queried when applying to health-related data that are combined with aviation-related data are:

- Which assets (data or service assets) are related to health category? (Listing 4)
- Which assets (data or service assets) are related to Coronavirus? (Listing 5)
- How many people recovered in Cyprus until 12-04-2020 from Coronavirus? (Listing 6)
- Which datasets can help me predict the Coronavirus transmission from incoming flights? (Listing 7)

```
SELECT
   (str(?id) AS ?asset_id) (?name AS ?asset_name)
   (str(?type) AS ?asset_type)
   (str(?categories) AS ?asset_categories)
WHERE {
   ?entity icarus:asset_ID ?id .
   ?entity icarus:asset_categories ?categories .
   ?entity icarus:asset_name ?name.
   ?entity icarus:asset_type ?type .
   filter(regex(?categories, "Health"))
}
```

Listing 4: SPARQL query for finding assets that are related to Health

```
SELECT
   (str(?id) AS ?asset_id) (?name AS ?asset_name)
   (str(?type) AS ?asset_type)
WHERE {
   ?entity icarus:asset_ID ?id .
   ?entity icarus:asset_name ?name .
   ?entity icarus:asset_description ?description .
   ?entity icarus:asset_type ?type .
   filter(regex(?name, "coronavirus", "i") ||
        regex(?description, "coronavirus", "i")) .
}
```

Listing 5: SPARQL query for finding assets that are related to Coronavirus

```
SELECT
  (group_concat(?rec; separator=" ") as ?recovered)
  (COUNT(*) AS ?successful_filters)
WHERE {
  ?dataset rdf:type icarus:Tabular_dataset .
  ?dataset icarus:has_column ?column .
  ?column icarus:has_type ?type .
  ?type rdf:type icarus:Disease .
  ?type icarus:row_id ?row .
  ?type icarus:disease_name | icarus:disease_country |
    icarus:disease_observation_date |
    icarus:disease_confirmed_recovered ?val .
  optional {?type icarus:disease_confirmed_recovered ?rec}
  filter (regex(?val, "coronavirus", "i") ||
    regex(?val, "cyprus", "i") || ?val > -1 ||
    ?val = "2020-04-12"^^xsd:date)
}
GROUP BY ?row
```

Listing 6: SPARQL query for finding the number of people that recovered from Coronavirus

```
SELECT
    (group_concat(?dname; separator=";") AS ?datasets_names)
    (group_concat(?dID; separator=";") AS ?datasets_ids)
WHERE {
    ?service rdf:type icarus:Service_Asset .
    ?service icarus:asset_ID ?serv_id .
    ?service icarus:asset_description ?serv_description .
    filter(regex(?serv_description, "Coronavirus", "i") &&
           regex(?serv_description, "transmission", "i") &&
           regex(?serv_description, "incoming", "i") &&
           regex(?serv_description, "flights", "i")) .
    ?service icarus:serviceContainsData ?dataset .
    ?dataset icarus:asset_name ?dname .
    ?dataset icarus:asset_ID ?dID .
}
GROUP BY ?serv_id
```

Listing 7: SPARQL query for finding datasets for the prediction of the Coronavirus transmission

4.4 Use Case Scenario: Application of the ICARUS Ontology to a Recommendation System

In this use case, the objective is to utilize the ICARUS ontology for providing high-quality recommendations of datasets and services to the users. As Figure 8 depicts, the system consists of the content-based and collaborative filtering components. In ICARUS, we use a weighted-based hybrid approach combining both models. In the following paragraphs, we focus on demonstrating the use of ICARUS ontology in both modules. The details of the ICARUS recommendation system are not given since it is out of the scope of this paper.

The content-based (*CB*) module generates recommendations by mapping the users' preferences, geolocation, and organization type with the respective information of the given datasets and services. Specifically, the ICARUS ontology would be used for capturing the structural and semantic characteristics of the various entities

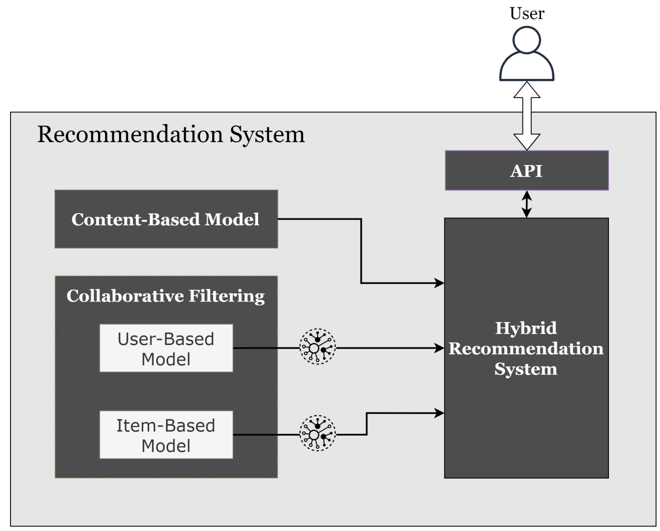

Figure 8: Recommendation System Architecture

involved in the given assets and facilitating the use of lightweight reasoning during the content-based recommendation process. Due to the semantic reasoning, we can reveal hidden relationships (e.g. inheritance) between entities, something impracticable using a relational database. Furthermore, by inferring additional relationships between users and assets, we can recommend assets that are connected indirectly to the user's preferences and needs. Listing 8 and 9 depict the SPARQL queries which are needed to retrieve the user's preferences and the asset's categories, respectively. In a nutshell, combining the outcomes of these queries, the recommendation model utilizes the semantic functionalities of the ICARUS Ontology and scores/ranks the most appropriate assets that best match a target user.

```
SELECT
    (str(?id) AS ?user_id)
    (str(?preferences) AS ?user_preferences)
WHERE {
    ?entity rdf:type icarus:Platform_User .
    ?entity icarus:platformUserID ?id .
    ?entity icarus:platformUserPreferences ?preferences .
}
```

Listing 8: SPARQL query for retrieving user-related data

```
SELECT
    (str(?id) AS ?asset_id) (?name AS ?asset_name)
    (str(?categories) AS ?asset_categories)
    (str(?type) AS ?asset_type)
WHERE {
    ?entity icarus:asset_ID ?id .
    ?entity icarus:asset_categories ?categories .
    ?entity icarus:asset_type ?type .
    ?entity icarus:asset_name ?name.
}
```

Listing 9: SPARQL query for retrieving the metadata of assets

On the other hand, collaborative filtering utilizes the ontology by capturing the interplay between organizations and assets to construct the interaction matrix. Specifically, the collaborative filtering component consists of two separate models: a user-based (UM) model and an item-based model (IM). The user-based model, suggests to a user, assets based on what assets similar users prefer. Alternatively, the item-based suggests assets that are similar to assets that users prefer. The success of the models is highly correlated with the input, as in most machine learning systems, and therefore the sparsity of the model determines the predictive power of the component. In this context, Listing 10 illustrates the SPARQL query in order to retrieve the exact interaction between an organization and an asset. Taking the exact interactions allows us to determine a different impact of each one of them. For example, we score lower a view interaction rather than a requested one. Therefore, we build a meaningful interaction matrix. Then, we utilize the KNN cluster-based machine learning algorithm to fill the gaps in it.

```
SELECT *
WHERE {
    ?user ?relationship ?asset
    FILTER (?relationship IN
        (icarus:userViewedAsset, icarus:userStarredAsset,
        icarus:userRequestedAsset, icarus:userPurchasedAsset)
    )
}
```

Listing 10: SPARQL query for retrieving relationships between users and assets

5 RELATED WORK

Ontological approaches are common in many domains such as in grid and cloud computing [33] [6], medicine [26], and information fusion [19]. Due to the massive growth of available aviation data, ontologies have also been popular for concept modeling in the aviation industry domain. They have been used to enrich the input datasets or data sources concerning the aviation domain, as well as other domains related to aviation such as transport, weather, health, etc. One of the most commonly used ontologies in the aviation domain is the NASA ATM (Air Traffic Management) Ontology [18]. It describes classes, properties, and relationships relevant to the domain of air traffic management and represents information pertinent to a broad and diverse set of interacting components in the US and the global airspace. Even though this ontology focuses on aviation, it is scoped sufficiently broadly to interconnect data from several different aviation realms, including flight, traffic management, aeronautical information, weather, and carrier operations.

The rest of the aviation ontologies, in contrast to the NASA ontology, represent only specific aspects of the aviation domain. For instance, the AIRCRAFT ontology [4] is a Web Ontology Language (OWL) ontology focuses on the physical structure of a typical passenger aircraft on a rather high level. Its main intention is to model standard fixed wing passenger aircrafts (e.g. Airbus A320, Boeing 747) on a basic level. In addition, in [31], an ontology has been constructed to integrate the knowledge for the design of aviation complex products. The collaborative design knowledge ontology of complex products ontology is an integration of multi-layer ontologies that represent the concepts related to the type of knowledge for the design of aviation products e.g. technical, characteristic, managerial knowledge. The ontology-based knowledge integration model consists of the resource layer, local ontology layer, global ontology layer, visualization layer and application layer.

Some other related works present ontologies from diverse domains that can also be related to aviation such as the Transport Disruption ontology [10], the QALL-ME ontology[14], the Ontology of Transportation Networks [20], and the Epidemiology Ontology (EPO) [24]. The Transport Disruption ontology models events related to travel and transport, which may disrupt the travel plans of an agent. The ontology was defined based on an analysis of information provided by transport authorities and operators of air public transport services, bus, rail and ferry. In particular, the ontology defines several subclasses of Event like type, location, time period, compound and causal relationships to other events, and any impact experienced by agents that have to adapt their plans because of it. The QALL-ME ontology [14] covers many aspects related to the tourism industry, including tourism destinations (i.e. cities and towns), tourism sites (i.e. accommodation, gastro, attraction, and infrastructure), tourism events (e.g. movie and show) and transportation. The Ontology of Transportation Networks (OTN) [20] models the most important aspects of traffic networks, transportation and locomotion, including classes such as airport area, urban area, industrial area, railways, ferries, parking, speed limits and many other. The EPO ontology describes several epidemiology and demography parameters as well as transmission of infection processes and participants. Its main purpose is to support the precise and comprehensive semantic annotation of epidemiology resources, such as documents, datasets, models and simulations. Finally, the OTN ontology is a direct encoding of the Geographic Data Files (GDF) standard in the OWL.

The main advantages of the ICARUS ontology, in contrast to the related ontologies can be summarized as follows: a) It facilitates the representation of an integrated model consisting of both general aviation concepts and of concepts from other diverse domains. To achieve this, the ICARUS ontology integrates concepts and relations from existing aviation-related domain ontologies such as the NASA ontology, and extend these ontologies with new concepts/relations. b) Due to its multi-layer design, the ICARUS ontology is easily extendable. For example, as shown in Section 4, it can be easily extended to represent data from different sources (e.g. Twitter data, epidemiological data). Also, all of the related ontologies either from aviation or diverse domains that have been discussed previously in this section, can be integrated and extend the ICARUS ontology as described in Section 3.3. c) By introducing the new class *Data_Value_type*, the ICARUS ontology is able to represent and map almost any data format of the row instances of the datasets such as timestamp, location, textual or numerical data.

6 CONCLUSION

In this paper, we present the ICARUS ontology, an aviation domain ontology designed using the multi-layer approach. The main strengths of the proposed ontology comparing to state of the art works, are its extendibility and interoperability due to the multilayer design, its ease of use on multiple aviation data sources of different format and structure such as purchased aviation datasets, epidemics datasets as well as data retrieved from Twitter. The proposed ontology enables the functionality of querying data for matching its concepts with dataset entries but also the functionality of answering competencies questions regarding the popularity of a particular airline or destination airport for a particular time period. Another application of the ICARUS ontology is to be used in combination with a recommender algorithm for recommending aviation datasets based on user's preferences.

Acknowledgement. This work is partially supported by the EU Commission in terms of ICARUS 780792 (H2020-ICT-2016-2017) project and by the Cyprus Research Promotion Foundation in terms of COMPLEMENTARY/0916/0010 project.

REFERENCES

[1] Building a knowledge graph for the air traffic management community. In *Companion Proceedings of The 2019 World Wide Web Conference*, WWW '19, page 700–704, New York, NY, USA, 2019. Association for Computing Machinery.

[2] Jose Aguilar, Marxjhony Jerez, and Taniana Rodríguez. Cameonto: Context awareness meta ontology modeling. *Applied Computing and Informatics*, 14(2):202 – 213, 2018.

[3] Grigoris Antoniou and Frank van Harmelen. *A Semantic Web Primer*. The MIT Press, 2004.

[4] Markus Ast, Martin Glas, Tobias Roehm, and VB Luftfahrt. *Creating an ontology for aircraft design*. Deutsche Gesellschaft für Luft-und Raumfahrt-Lilienthal-Oberth eV, 2014.

[5] Ghislain Auguste Atemezing and Florence Amardeilh. Benchmarking commercial rdf stores with publications office dataset. In *European Semantic Web Conference*, pages 379–394. Springer, 2018.

[6] Nick Bassiliades, Moisis Symeonidis, Panagiotis Gouvas, Efstratios Kontopoulos, Georgios Meditskos, and Ioannis Vlahavas. Paasport semantic model: An ontology for a platform-as-a-service semantically interoperable marketplace. *Data & Knowledge Engineering*, 113:81 – 115, 2018.

[7] ICARUS Consortium. D3.3–Architecture, Core Data and Value Added Services Bundles Specifications. (last accessed: 03.03.2020).

[8] ICARUS Consortium. ICARUS. Aviation-driven Data Value Chain for Diversified Global and Local Operations, 2020. (last accessed: 19.02.2020).

[9] ICARUS Consortium. D3.1–ICARUS Architecture, APIs Specifications and Technical and User Requirements, 31 March 2019. (last accessed: 03.03.2020).

[10] David Corsar, Milan Markovic, Peter Edwards, and John D Nelson. The transport disruption ontology. In *International Semantic Web Conference*, pages 329–336. Springer, 2015.

[11] Liviu-Adrian Cotfas, Camelia Delcea, Ioan Roxin, and Ramona Paun. Twitter Ontology-Driven Sentiment Analysis. In Dariusz Barbucha, Ngoc Thanh Nguyen, and John Batubara, editors, *New Trends in Intelligent Information and Database Systems*, pages 131–139, Cham, 2015. Springer International Publishing.

[12] Orri Erling and Ivan Mikhailov. Rdf support in the virtuoso dbms. In *Networked Knowledge-Networked Media*, pages 7–24. Springer, 2009.

[13] Orri Erling and Ivan Mikhailov. Virtuoso: Rdf support in a native rdbms. In *Semantic Web Information Management*, pages 501–519. Springer, 2010.

[14] Óscar Ferrández, Christian Spurk, Milen Kouylekov, Iustin Dornescu, Sergio Ferrández, Matteo Negri, Rubén Izquierdo, David Tomás, Constantin Orasan, Günter Neumann, Bernardo Magnini, and José Luis Vicedo González. The qall-me framework: A specifiable-domain multilingual question answering architecture. *J. Web Semant.*, 9:137–145, 2011.

[15] Rakebul Hasan and Fabien Gandon. A machine learning approach to sparql query performance prediction. In *2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)*, volume 1, pages 266–273. IEEE, 2014.

[16] Pascal Hitzler, Aldo Gangemi, and Krzysztof Janowicz. *Ontology engineering with ontology design patterns: Foundations and applications*, volume 25. IOS Press, 2016.

[17] Richard M Keller. Ontologies for aviation data management. In *2016 IEEE/AIAA 35th Digital Avionics Systems Conference (DASC)*, pages 1–9. IEEE, 2016.

[18] Richard M Keller. The NASA air traffic management ontology: Technical documentation. 2017.

[19] Mieczyslaw M. Kokar, Christopher J. Matheus, Kenneth Baclawski, Jerzy Letkowski, Michael L. Hinman, and John J. Salerno. Use cases for ontologies in information fusion. 2006.

[20] Bernhard Lorenz, Hans Jürgen Ohlbach, and Laibing Yang. Ontology of transportation networks. 2005.

[21] Fadi Maali, John Erickson, and Phil Archer. Data catalog vocabulary (dcat). https://www.w3.org/TR/vocab-dcat/. Accessed: 2019-06-20.

[22] Mohamed Morsey, Jens Lehmann, Sören Auer, and Axel-Cyrille Ngonga Ngomo. Dbpedia sparql benchmark–performance assessment with real queries on real data. In *International semantic web conference*, pages 454–469. Springer, 2011.

[23] Mark A. Musen. The protégé project: a look back and a look forward. *AI Matters*, 1(4):4–12, 2015.

[24] Catia Pesquita, João D Ferreira, Francisco M Couto, and Mário J Silva. The epidemiology ontology: an ontology for the semantic annotation of epidemiological resources. *Journal of biomedical semantics*, 5(1):4, 2014.

[25] Lynn M Schriml, Elvira Mitraka, James Munro, Becky Tauber, Mike Schor, Lance Nickle, Victor Felix, Linda Jeng, Cynthia Bearer, Richard Lichenstein, Katharine Bisordi, Nicole Campion, Brooke Hyman, David Kurland, Connor Patrick Oates, Siobhan Kibbey, Poorna Sreekumar, Chris Le, Michelle Giglio, and Carol Greene. Human Disease Ontology 2018 update: classification, content and workflow expansion. *Nucleic Acids Research*, 47(D1):D955–D962, 11 2018.

[26] Barry Smith and Werner Ceusters. Towards industrial strength philosophy: how analytical ontology can help medical informatics. *Interdisciplinary Science Reviews*, 28(2):106–111, 2003.

[27] Grimm Stephan, Hitzler Pascal, and Abecker Andreas. Knowledge representation and ontologies. *Semantic Web Services: Concepts, Technologies, and Applications*, pages 51–105, 2007.

[28] Mari Carmen Suárez-Figueroa, Asunción Gómez-Pérez, Enrico Motta, and Aldo Gangemi. *Ontology Engineering in a Networked World*. Springer, 2012.

[29] Audun Vennesland, Richard M. Keller, Christoph G. Schütz, Eduard Gringinger, and Bernd Neumayr. Matching ontologies for air traffic management: a comparison and reference alignment of the airm and nasa atm ontologies. In *OM@ISWC*, 2019.

[30] Martin Voigt, Annett Mitschick, and Jonas Schulz. Yet another triple store benchmark? practical experiences with real-world data. In *SDA*, pages 85–94. Citeseer, 2012.

[31] Y. Wang, L. Song, F. Wang, H. Ji, and X. Dai. Ontology-based knowledge integration of collaborative design for aviation complex products. In *2011 IEEE 18th International Conference on Industrial Engineering and Engineering Management*, volume Part 3, pages 1894–1898, Sep. 2011.

[32] Wei Xing, Oscar Corcho, Carole Goble, and Marios D. Dikaiakos. An acton-based semantic information service for grids. *Future Generation Computer Systems*, 26(3):324 – 336, 2010.

[33] Wei Xing, Marios D Dikaiakos, and Rizos Sakellariou. A core grid ontology for the semantic grid. In *Sixth IEEE International Symposium on Cluster Computing and the Grid (CCGRID'06)*, volume 1, pages 178–184. IEEE, 2006.