Ontologies for Aviation Data Management

Richard M. Keller
Intelligent Systems Division
NASA Ames Research Center
Moffett Field, CA, USA
rich.keller@nasa.gov

Abstract—**Managing complex aviation data can be a significant challenge for any enterprise – whether a government agency, airline, airframe manufacturer, or aviation service provider. To handle this challenge, data models are typically developed to characterize and manage the data generated, used, and stored by a given enterprise. Unfortunately, different data providers employ qualitatively different data models, and this gives rise to problems exchanging data across organizational boundaries. Over the past decade, these problems have motivated data producers and consumers to look toward standardized data exchange models to address data interoperability. In this paper we examine some of these standardized data exchange models and compare them with a new type of data model based on ontologies. Ontology models have emerged in recent years from a confluence of research in the artificial intelligence, semantic web, and information science communities. This paper introduces ontology models, provides several use cases for ontologies relevant to aviation data management, and summarizes state-of-the-art aviation prototype applications that utilize ontologies.**

Keywords—aviation data management; data models; data interoperability; ontologies

I. INTRODUCTION

The volume of aviation data generated by industry and government sources has grown steadily over the past decade as an increasing number of data providers have begun publishing their data through various digital distribution methods. This trend has been a boon to those in the industry who use data to improve aviation systems, derive cost savings, enhance safety, and analyze bottlenecks. But along with additional data comes increased data management complexity.

Managing data from multiple sources in a consistent and coherent fashion can be a significant challenge for any enterprise – whether a government agency, airline, airframe manufacturer, or aviation service provider. Typically data models are developed to manage the data generated, used, and stored by a given enterprise [1]. A data model encodes the structure, format, and constraints on stored data values, as well as their relationship(s) to real-world entities that give rise to the data. Unfortunately, aviation data providers can employ quite different data models, and these data models can vary along various dimensions:

- Data encoding format: Providers can use different formats to encode data for storage and transmission. For example, an airline carrier ID field could be stored in a variety of ways: as a character field containing a three letter International Air Transport Association (IATA) code or a four letter International Civil Aviation Organization (ICAO) code, or as an integer corresponding to a numeric key into a proprietary air carrier database table, for instance.

- Data field naming: Data values are identified by field names to label their content. However, without standardization, field names assigned to values can be misleading and can lead to confusion. For the value of an 'aircraft departure time' field, one provider's data model may use the name 'dptTime' while another may use 'DT'. Reconciling these differences across models can be challenging.

- Data semantics: Even if two data fields are identically named across providers, that doesn't ensure the data represents the same information. The actual meaning of 'aircraft departure time' across two providers may be very different: it may correspond to a filed vs. a scheduled vs. an actual departure time from either the gate or the runway. It would be difficult to determine whether the fields in two data models are identical without more specific contextual information.

- Spatial and temporal resolution: Data values for a given parameter may be recorded at different temporal frequencies (e.g., once per minute vs. once per second) and aggregated across different time periods (e.g., hourly, yearly) or spatial regions (e.g. across airspace sectors, geographic regions, or political entities). Without conveying knowledge of the spatial or temporal extent, data cannot be properly interpreted or combined with other data.

- Measurement unit conventions: Units are often omitted and left implicit in data storage schemes. However this can lead to problems when different units are employed across different systems. Metric vs. imperial is a well-known example, but also consider how altitudes may be stored according to feet (20000) vs. flight level (200).

- Data organization principles: Each data provider has a specific world view that makes sense for its business, and this unique perspective guides the way in which the provider's data is structured, stored, and transmitted. But this world view may turn out to be parochial and difficult to interpret by the recipient of the data. For

example, an aircraft engine manufacturer may structure and transmit its engine data as a series of separate tables, each describing one aspect of the data, such as technical specifications, performance characteristics, pricing, maintenance, etc. Each table includes rows containing data values corresponding to different engine models. However, the consumer of this data may be interested in a just a single aircraft engine model. In this case, the manufacturer's multi-table partitioning makes it cumbersome to fetch and assemble the desired data.

With all these potential differences in data models, it can be extremely challenging to work with data from multiple providers. Responding to this problem, the aviation industry has developed a number of standardized data models to facilitate data exchange and promote interoperability. This paper explores an alternative data management approach that uses highly-composable, semantics-rich ontologies to describe data and mediate disparities across different data sources. We give a brief introduction to ontologies and describe how they are being used in aviation; then we illustrate how ontology models can support a rich variety of aviation data management use cases and describe current aviation industry prototype applications that utilize ontologies. But first, we start with an examination of current industry approaches to standardization of data exchange.

II. STANDARDIZED DATA EXCHANGE MODELS

The European Organization for Safety of Air Navigation (EUROCONTROL), the US Federal Aviation Administration (FAA), and IATA have taken the lead in developing industry-wide standards for data interoperability. These organizations are developing standardized languages for the transmission of data between producer and consumer. As part of their efforts, they have developed a variety of data models. EUROCONTROL and FAA have co-developed Flight Information Exchange Model (FIXM), Aeronautical Information Exchange Model (AIXM), and Weather Information Exchange Model (WXXM). FIXM [2, 3] captures data related to flights throughout their lifecycle from pre-flight planning through departure, en-route, and arrival segments, including data on aircraft location, aircraft characteristics, and cargo contents. AIXM [3-5] describes airspace routes, procedures, boundaries, fixes, navaids, and airport surface elements (runways, gates, etc.). WXXM [3, 6] models meteorological observations, forecasts, and measurement procedures for weather features and phenomena with spatial and temporal extent. Independent from governmental standardization efforts, the airline industry has been developing Airline Industry Data Model (AIDM) [7]. AIDM, which is being spearheaded by IATA, overlaps with portions of AIXM and FIXM, but also covers other areas specific to the concerns of air carriers and airports, such as fuel, baggage, reservations, ticketing, and ground servicing. In addition to these key models, more specialized data models are being developed (in some cases as extensions to the aforementioned): Maintenance Management Exchange Model (MMIXM) [8], which covers maintenance logging, event coordination, and asset tracking; Aerodrome Mapping Exchange Model (AMXM) [9] for interchanging airport map information; Traffic Management Information Exchange Model (TMXM) [10] for standardizing information about traffic management initiatives; and IWXXM[11], a variant of WXXM developed by ICAO and the World Meteorology Organization (WMO) and focused more narrowly on exchange of specific weather products standardized by ICAO and WMO.

While these new standardized data models represent progress in advancing aviation data interoperability, there are some challenges that still remain. The first involves scoping. Each of these models is scoped narrowly by design to cover a circumscribed aviation subdomain. Aviation data, however, knows no bounds, and may not fit neatly within the scope of a single model. For example, a flight whose data can be described by FIXM, traverses the airspace using routes and procedures defined in AIXM. In fact, there are many points of overlap between AIXM and FIXM concepts, but those overlaps are not well articulated. Concepts in common across these models may be named differently and are generally expressed at different levels of detail.

There is widespread agreement that AIXM, FIXM, and WXXM, which were developed by three different communities, must be 'harmonized' and made interoperable. However, the authors of the FIXM Model Construction Guidelines [12] provide this assessment of the recent state of interoperability across the exchange models: "It is recognized that the harmonization of AIXM, FIXM and WXXM is a desirable goal. However, at the time of writing [December 2014] there has been little progress towards achieving that harmonization." EUROCONTROL's Single European Sky ATM Research (SESAR) organization is making progress on the mechanics of harmonization by developing cross-vocabulary correspondence mechanisms using their ATM Information Reference Model (AIRM) [13]. The AIRM serves as EUROCONTROL's master data dictionary of ATM entites, properties, and relationships. But the actual work of mapping exchange model terms to AIRM terms still lies ahead.

In the meanwhile, data products with contents that stray across the existing artificial model boundaries are not fully supported. Questions remain as to how and whether these different models can be fit together to build a more comprehensive and extensible common aviation data model. Without a broader aviation data model, it will not be possible to integrate data from multiple sources to enable cross-domain aviation applications. Consider an airspace analysis application that examines historical weather, flight, and air traffic patterns to uncover potential methods of increasing throughput in the air traffic control system. Such a system would need to access integrated data across all three of these standard data models.

Another concern with respect to the standard models is 'legacy use bias'. These models were developed primarily to facilitate exchange of preexisting aviation data products, and can be biased to encode data in ways that are historically consistent with those products. As a result, there is a concern that these models have not been sufficiently generalized to serve future products or different purposes. The model development guidelines for FIXM explicitly address this concern [12], but this type of inherent bias can be difficult to eliminate.

A final concerns relates to agility and maintainability of the standard models. AIXM, FIXM, WXXM, and AIDM were each defined using the Unified Modeling Language (UML) [14] as the basis for their specification. UML is a diagrammatic design formalism for specifying software systems, but is also used for data modeling. With UML's constructs for software object modeling, a model that describes the data associated with aviation business domain entities, their properties, and interrelationships can be defined. But the use of UML makes data model changes more cumbersome and less agile; this is because UML serves only as a data model specification language and must be mapped to a target data representation language – typically XML – for implementation. Keeping the data model and the data representation synchronized in the face of changes to the model can be a real logistical challenge and is a point of brittleness in the UML approach.

III. WHAT ARE ONTOLOGIES?

An ontology [15-17] is a type of data model that has emerged in recent years from a convergence of research in the artificial intelligence (AI), semantic web, and information science communities. Ontologies were originally developed as a type of AI knowledge representation formalism intended to represent information stored in large knowledge bases [18] and on web pages [19] for processing by intelligent agents performing reasoning and inference tasks. Although the uses of ontologies go beyond data modeling, they have some properties that can provide additional value over traditional data modeling languages.

Ontologies are based on simple subject-predicate-object statements called 'triples', which can be expressed in a language called Resource Description Framework (RDF) [20]. RDF, along with RDF Schema (RDFS) [21, 22], form the basis for the Web Ontology Language (OWL) [23, 24]. OWL is the principal ontology language in use today. Triples are statements that describe entities in the modeling domain, their properties, and their relationships to other entities. The subject in a triple represents the data entity being described; the predicate expresses a relationship between the subject data entity and the object, which can be either another data entity or a literal property value associated with the subject. Here are some sample triples describing flight UAL269, along with an explanation for each triple:

$$<\text{UAL269 hasOriginAirport KJFK}> \quad (1)$$

The triple in (1) expresses the fact that *UAL269* is an entity in the domain and is related to the entity *KJFK* through the predicate *hasOriginAirport*.

$$<\text{UAL269 hasDestinationAirport KSFO}> \quad (2)$$

In (2), we see that *UAL269* is also related to the *KSFO* entity by the *hasDestinationAirport* predicate.

$$<\text{UAL269 departsRunwayAt 2016/05/25@10:32GMT}> \quad (3)$$

In (3) the *departsRunwayAt* predicate links *UAL269* to its property value, not to another entity; the property value is a time literal.

$$<\text{UAL269 hasCarrier UnitedAirlines}> \quad (4)$$

In (4) *UnitedAirlines* is another entity in the domain.

$$<\text{UnitedAirlines hasICAOcode “UAL”}> \quad (5)$$

In (5) *UnitedAirlines* has an associated property value "UAL", a string which corresponds to its ICAO code.

$$<\text{UAL269 type Flight}> \quad (6)$$

In (6) the special predefined predicate *type* is used to indicate that the subject is an instance of the object entity; UAL269 is a type of Flight.

$$<\text{hasCarrier type Predicate}> \quad (7)$$

We can also make statements about predicates; in (7) we state that hasCarrier is an instance of a *Predicate*.

$$< \text{hasCarrier domain Flight}> \quad (8)$$

In (8) we state that the *domain*[1] of the *hasCarrier* predicate is an instance of a *Flight*.

$$< \text{hasCarrier range Airline}> \quad (9)$$

In (9) the *range*[1] of the *hasCarrier* predicate is an instance of an *Airline*.

Ontologies can be visualized best as directed graph structures encoding triples. Each node in the graph represents a subject or an object, and each directed link represents a predicate. Figure 1 depicts a fragment of an ontology that contains the sample triples describing UAL269.
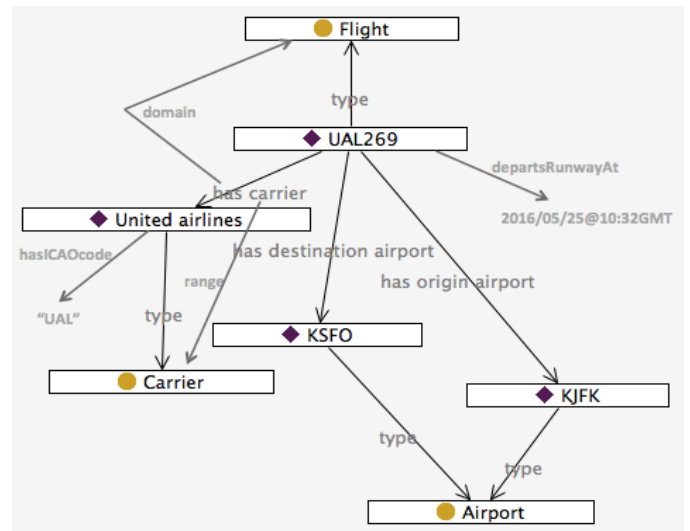

Figure 1: Graphical representation of sample triples

---

[1] The terms 'domain' and 'range' draw from a correspondence between a mathematical function and a predicate; a function transforms a value in its domain into a value in its range.

There are a few key points of differentiation and advantage for ontologies versus conventional data models:

- Ontologies can serve simultaneously as a specification language and an implementation language. In other words, an ontology language can describe data objects, attributes, constraints, and relationships at an abstract level, and can also represent actual data instances. This is evident in the set of triples shown above. There are statements describing an actual instance of a flight (triples (1)-(4) and (6)), as well as statements about how flights in general can be related to Airlines in general (triples (7)-(9)). This uniformity and reflexivity of representation makes it easy to reason about data and metadata in one framework. This is not the case with the UML/XML approach or with relational database languages such as SQL, where the database schema definition information is represented separately from the data in tables. With ontologies, the data and the metadata, both in the form of triples, are maintained in a 'triple store' [25], which is a specialized data repository for storing, retrieving, and inferencing with triples.

- The fine-grained modularity of the triple representation makes incremental changes in data models relatively simple, and isolates the impact of changes to localized modifications. Rather than bundling many modeling statements together in a complex modeling language construct, each triple is independent and can be incrementally added or deleted from the data model as needed. In contrast, UML packages data properties and relationships into more structured, less flexible representations that are not as easily modified.

- Ontologies are built soundly upon principles of formal logic, which can be used to reason unambiguously about the data and draw inferences based on logic rules. These formal underpinnings provide a means of extending the data beyond what is explicitly stored in a triple store and permits the generation of implicit data from existing data via inference. In addition, rule formalisms have been built on top of ontologies (e.g., SPIN [26], SWRL [27]) to extend the native logic capabilities further, and augment data using if-then style rules.

- RDF and OWL-based ontologies were designed and developed for the distributed environment of the Web. Numerous ontology repositories have been deployed to archive reusable ontology models in a variety of domains [28, 29]. The community practice is to develop and publish modular, general-purpose ontologies that can be imported and augmented to build up larger customized ontologies.

- The data and metadata stored in an ontology can be retrieved using a special-purpose standardized graph query language called SPARQL [23]. This is analogous to SQL for relational databases. However, unlike SQL, SPARQL queries can range over either metadata or data, because they are both stored in the same triple format representation.

These advantages over conventional data models make ontologies an attractive alternative and have encouraged some in the aviation data management community to experiment with ontologies in a variety of settings and applications. We describe some of these in the next section.

IV. ONTOLOGY-BASED AVIATION APPLICATIONS

Prior to 2015, there was a scattering of published work on ontology-based aviation applications [30-32], but there had been no attempt to convene a dedicated forum on this topic. In 2015, an invitational workshop was held to bring together a community of aviation data management professionals focused in this area. The Semantic Web for Air Transportation (SWAT) workshop [33] brought together 25 international participants from government agencies (FAA, NASA, EUROCONTROL, Department of Transportation (DOT) / Volpe, Department of Defense (DOD), Air Services Australia), standards bodies (OGC – Open Geospatial Consortium, ICAO), university labs (MIT Lincoln Labs), and industry (Mosaic ATM, Frequentis, Securboration, Alion, Booz-Allen-Hamilton). The purpose of the meeting was to exchange information about ongoing projects, use cases, and shareable community resources such as ontologies, development tools, and deployment strategies. Presented at the workshop were a rich variety of aviation data management application use cases covering data integration from heterogeneous sources, data exchange across software systems, software interoperation, data search and retrieval, data extraction, and data publication. The following subsections highlight some of the ontology applications presented at the workshop.

A. NASA Air Traffic Management Data Integration [34-36]:

The National Aeronautics and Space Administration (NASA) has developed a prototype data integration system that demonstrates how ontologies can be used to integrate, query, and search over various sources of heterogeneous air traffic management (ATM) data, including data from FAA, National Oceanic and Atmospheric Administration (NOAA), NASA, and other providers. In this prototype, a common ontology is used to bridge multiple types of aviation data models and enable cross-dataset querying. In general, cross-dataset querying is very challenging due to differing data formats, nomenclature, and organizational structure, as discussed in Section I. Data can only be combined from multiple sources by expending significant effort to write customized code that integrates selected data on an as-needed, piecemeal basis. To address this problem, NASA is building an integrated data source that obviates this need to write special-purpose, one-off integration code.

As part of this approach, an overarching data model (the NASA ATM Ontology) has been designed and implemented to serve as a backbone upon which to overlay data from multiple sources. The ontology data model is scoped sufficiently broadly to interconnect data from several different aviation realms, including flight, traffic management, aeronautical information, weather, and carrier operations. Representative entities that are modeled within the ontology include the following:

- Flights: including flight plans and radar flight tracks;

- Aircraft and manufacturers: including aircraft characteristics, models, and makes;

- Airports and infrastructure: runways, taxiways, terminals, gates, and deicing pads;

- Airlines;

- US National Airspace System (NAS) facilities: e.g., the Air Traffic Control System Command Center (ATCSCC), Enroute Air Traffic Control Centers (ARTCCs), Terminal Radar Approach Control Centers (TRACONs), and airport towers;

- Air traffic management initiatives (TMIs): e.g., Ground Delay Programs (GDPs), Ground Stops (GSs), ReRoutes, and Miles-In-Trail restrictions (MIT);

- Surface weather conditions and forecasts: including meteorological conditions, Meteorological Terminal Aviation Routine Weather (METAR) reports, and Terminal Area Forecast (TAF) reports;

- Airspace components: sectors, fixes, routes, and airways;

- Departure/arrival routes

The ontology is currently populated with instance data corresponding to over 100K flights arriving and departing the three major airports in the New York Metroplex (KEWR, KJFK, KLGA) during July 2014. This data is encoded in approximately 250M triple statements and incorporates the following sources: flight track data from FAA's ASDI (Aircraft Situation Display to Industry) data feed; airport weather from the NOAA's METAR and TAF data feeds; information on traffic management initiatives from FAA's Command Center website; historical traffic counts and delay statistics from the DOT's Aviation System Performance Metrics (ASPM) database; infrastructure data on US routes, fixes, airways, and sectors used by FAA's En Route Automation Modernization (ERAM) system; aircraft information from FAA's aircraft registry and from the Commercial Aviation Safety Team (CAST)/ICAO common aircraft taxonomy [37]; and web-based information on airlines, airport terminals and gates. Of the 250M triples, about 99.5% represent specific flight, weather, or traffic management data. These are dominated overwhelmingly by triples capturing flight track parameters from ASDI – such as altitude, location, and groundspeed – captured at the rate of one per minute of flight. The remaining 0.5% of triples capture mostly static aeronautical infrastructure information (e.g., data on airspace routes, sectors, and fixes, NAS facilities, airports, aircraft makes and models, airlines, etc.).

B. TMI Attribute Standardization (TAS) (FAA & Mosaic ATM)[38]:

TAS was initiated to address data standardization issues in the issuance and tracking of TMIs across FAA systems. TMIs are issued to manage traffic imbalances in the NAS [39]. TMIs are managed and/or referenced by numerous FAA systems, including National Traffic Management Log (NTML), Information Display System (IDS), Traffic Flow Management System (TFMS), and Flight Schedule Monitor (FSM). But these systems lack a standardized vocabulary or a common conceptual viewpoint on TMIs. As a result, interoperation across systems is difficult and can lead to inconsistent reporting of TMIs in these different systems. A TMI Ontology was built as part of the TAS effort to define a set of common concepts and relationships pertinent to the issuance of TMIs. TMIs comprehensively modeled in this ontology include GDPs, GSs, MITs, Airspace Flow Programs, Speed Restrictions, and many others. The ontology links to NAS concepts from the AIXM model (described in Section II) that are relevant to the specification of TMIs (e.g., airways, routes flow controlled areas), as well as relevant concepts from FIXM (e.g., carriers and flights impacted by TMIs) and weather-related causes from WXXM.

C. FAA Enterprise Information Management [40]:

The FAA Enterprise Information Management organization is charged with improving access to internal FAA information across the Agency. As part of this effort, FAA has developed an ontology to standardize terminology used in FAA internal documents and thereby facilitate document search, document discovery, and ultimately, improved decision-making. The FAA has experimented with automated metadata tagging capabilities to build structure on top of free-form and semi-structured text documents. The basis for these efforts is a thesaurus of air transportation terms in combination with an Air Transportation Information ontology which serves as an upper ontology for the transportation domain. (An upper ontology provides a broad conceptual framework under which more specific concepts can be structured [41-43].) The ontology was developed using a methodology based on the notion of interrogatories [44] and was applied to capture connections among concepts in the transportation domain. The ontology focuses on concepts relevant to FAA's organizational functions, and includes the following broad categories of entities: people with their roles and organizations; air transportation activities (design, planning, oversight, assessment, enforcement); resources (financial, human, information, physical); transportation conveyances (aircraft, landcraft, spacecraft, watercraft); aircraft systems; rules, procedures, and legislative mandates; and others.

D. FAA Web Service Description Ontological Model [45]

System Wide Information Management (SWIM) [3] is a data distribution framework being implemented by FAA and EUROCONTROL to standardize mechanisms for delivery of ATM data between producers and consumers using a subscription-based service oriented architecture. As part of the SWIM architecture, data providers create services to access their data. For the FAA, these services are published in the NAS Services Registry/Repository (NSRR) [46]. NSSR is a catalog of all SWIM services and provides documentation on various aspects of each service, including its provider, functionality, quality characteristics, interface, and implementation. As part of this effort, FAA has developed the Web Service Description Ontological Model (WSDOM) to standardize information and metadata pertinent to describing

SWIM services and to facilitate interchange of service data between service providers and the FAA. The intent behind the ontology is to make service definitions clear, unambiguous, and discoverable by both humans and computer systems. WSDOM consists of ontology classes covering the key notions of service profile, service interface, service implementation, stakeholder, and document. WSDOM is patterned after the OWL-S semantic web services description ontology [47].

E. SWIM Documentation Controlled Vocabulary (FAA) [48]

FAA has developed a SKOS-based controlled vocabulary (CV) that provides a single source for terms and definitions used in SWIM-related documentation. SKOS (Simple Knowledge Organization System) is a W3C standardized RDF ontology for representing controlled vocabularies, thesauri, and taxonomies [49]. The SWIM CV includes textual definitions of SWIM-related terms and their connections to broader, narrower, and otherwise related terms within the CV. For example, the CV includes the term 'message-oriented middleware' defined as "Software or hardware infrastructure supporting sending and receiving messages between distributed systems." The term JMS (Java Messaging Service) is defined as a narrower term meaning "A Java-based application programming interface (API) that provides a common way for Java programs to create, send, receive, and read an enterprise messaging system's messages." SWIM documentation and web pages can reference term definitions in the CV using web Universal Resource Identifiers (URIs).

F. Semantic NOTAMs (Frequentis & University of Linz) [30, 50, 51]

Notices to Airmen (NOTAMs) are advisories to pilots concerning establishment of, or changes to flight procedures, airport facilities, flight services, or other potential operational hazards. As part of the flight planning and pre-flight briefing process, pilots must review all of the NOTAMs pertinent to their filed route of flight. This is a tedious process and there is serious pilot information overload due to the sheer number of NOTAMs that are potentially relevant to a flight – perhaps 50 pages of NOTAMs for a domestic US flight. The Semantic NOTAM (SemNOTAM) system annotates NOTAMs with temporal, spatial, equipage, and operational conditions under which a NOTAM is applicable. These annotations are expressed using terms and relationships defined in the SemNOTAM ontology, which is derived from the AIXM model. The system attempts to narrow down the pilot briefing package using rules that examine NOTAM annotations and filter out NOTAMs that are irrelevant to a specific flight. For example, a NOTAM may not be relevant to a flight based on its equipment, routing, and origin/destination airports. SemNOTAM also enables search and grouping of NOTAMs based on their annotations.

G. Cross Organization Semantic Services (Alion, Securboration) [52]

The goal of the Cross Organization Semantic Services (CrOSS) project was to provide tools that harvest, organize, enhance, and analyze information from textual documents, and make the information readily available to support decision-makers in various domains. One of the domains where CrOSS has been applied is the analysis of FAA rules and regulations pertaining to the operation of Unmanned Aeronautical Systems (UASs). In this domain, the input documents include the Federal Aviation Regulations (FARs), the FAA Aeronautical Information Manual (AIM), and FAA Advisory Circulars. The documents are processed to identify textual terms relevant to regulatory concerns. These terms are identified by matching the text against a UAS 'topic ontology'. The topic ontology contains concepts related to UASs and regulatory processes (e.g., aviation hazards, pilot communications, avoidance maneuvers) as well as numerous lexical forms associated with those concepts. In this sense, the ontology serves as both a knowledge model and a thesaurus or lexicon. The document is annotated with matching ontology concepts, and the ontology can then be queried to retrieve documents matching desired UAS concepts.

V. USE CASES

Drawing from the limited set of projects presented at the SWAT workshop, discussions with aviation data management practitioners, and transferrable experiences in the ontology development community, this section describes a set of different use cases for ontologies in aviation.

A. Software system interoperation

Ontologies can be used to facilitate communication and coordination among software systems with differing programmatic interfaces. An ontology can serve as a lingua franca for translating between differing systems. For example, different air traffic simulation systems typically utilize differing input and output parameters and even different temporal or spatial structures for describing the progress of aircraft through the NAS. These differences impede the sharing of information and simulation results across simulation systems. Ontologies can provide a common interchange language for mediating the exchange of information across systems. Another example comes from the TAS project described in Section IV.B. Multiple FAA systems contain information about TMIs, but the information is not interchangeable due to the use of subtly differing models of TMIs in the systems' software. By developing a standardized conceptual model of TMI information, software engineering costs can be decreased, data quality enhanced, and information can be interchanged across system boundaries.

B. Integration of information from multiple sources

Semantic data integration [53, 54] is a key use case for ontologies. An ontology can serve as the scaffolding upon which to overlay information from heterogeneous sources to form a single integrated data source with consistent spatial, temporal, and conceptual underpinnings. This is the approach taken by the NASA ATM Data Integration project described in Section IV.A. In that project, heterogeneous information about aircraft flight tracks, en-route and surface weather, and TMIs – expressed in varying data formats, based upon differing conceptual data models, and produced by different data providers – are integrated using an ontology.

C. Data query and search

Because a triple store both describes a data model and holds actual data and interrelationships, it can serve as a type of interconnected database to be used in query, search, and analysis applications. For example, NAS stakeholders (FAA, airlines, aviation industry, aviation researchers) have questions about operations that cannot easily be resolved without performing expensive and time-consuming data analyses requiring custom data acquisition, processing, and integration. For instance, to answer the following question requires data from many disjointed sources: "During the 2009-2012 winter seasons, locate all United Airlines flights that were rerouted due to weather after departing Chicago-O'Hare under freezing ground temperatures after ground delays of over 60 minutes." Data integrated and organized using an ontology can serve as a comprehensive centralized database capable of answering this type of cross-cutting question.

D. Terminology standardization

One common use of an ontology is to standardize terminology. This was seen in the SWIM Documentation Controlled Vocabulary project described in Section IV.E. Another example comes from attempts to standardize the description of aircraft make, model, and series. FAA and CAST/ICAO [37] use different and sometimes conflicting vocabularies in their separate standardization efforts. This makes it virtually impossible to compare and combine essential aircraft characteristics (e.g., max takeoff weight, wingspan, engine type) across FAA and CAST/ICAO sources, neither of which is comprehensive by itself. Ontologies can provide an intermediary terminology that enables translation from one vocabulary to another, or can serve as a basis for creating a new standard for adoption.

E. Data inference

By using rules and formal properties associated with ontologies, existing data can be checked for validity and new data can be generated from existing data. For instance, the concept of a 'weather-delayed flight' could be defined within an ontology (i.e., expressed in terms of a constraint involving scheduled time, actual time, weather, airport surface conditions, etc.) and then flights could be automatically classified by the ontology as to whether they fit this definition or not. This type of automatic classification, called subsumption [55], is built into the formal underpinnings of ontologies. In a similar vein, data validity constraints can be defined in an ontology, and data falling outside of allowed values can be automatically classified as erroneous data.

F. Data publishing

Ontologies play a key role in Linked Open Data (LOD). LOD is a set of technologies and best practices for publishing, interconnecting, and searching structured, publically available data on the web [56, 57]. Data from ontologies is typically stored in triple stores or other data repositories that can be exposed through SPARQL endpoints, which are web servers that can accept SPARQL queries and return ontology data in response to external requests. LOD connects and aggregates data from multiple SPARQL endpoints. As an example of how LOD could be used for aviation data, consider that NASA aeronautics researchers generate a considerable amount of data as part of their research activities, including output from analyses, studies, and simulation runs. Unfortunately there are no standards for sharing aeronautics data within NASA, with other government agencies, or with the public, despite mandates for increased sharing of data produced by federally funded research [58]. Ontologies can play a role in providing a standardized markup language for data produced by government researchers, and in publishing that data in triple stores connected to SPARQL endpoints. If all researchers annotated their data with machine-interpretable, self-describing semantic metadata markup, it would be easier to combine data across research studies and to share data with the public.

G. Decision support

It is worth noting that several of the ontology use cases described at the SWAT workshop ultimately involve decision support of one kind or another, providing assistance to analysts, policy-makers, regulators, air traffic managers, and operations personnel. Here are some examples of decision support tasks facilitated using ontologies and reported at SWAT:

- Regulatory analysis and compliance monitoring;
- Aviation software safety case validation [59];
- Post-incident support for analysis of safety, maintenance, or operational events;
- Pre-flight and in-flight pilot briefings;
- Airspace operations bottleneck analysis

H. Information extraction

One final use case for ontologies is the extraction of semantic information from unstructured text documents. This is accomplished by matching terms and relationships parsed from natural language text against terms and relationships in an ontology. If a match is found, the ontology information can be used to augment or amplify information in the text. The ontology can also serve as a structured index into the textual corpus. In particular, SPARQL queries can be used to retrieve textual passages relevant to concepts in the ontology. Both the FAA Enterprise Information Management project (Section IV.C) and the CrOSS project (Section IV.G) annotate FAA internal documents and regulations in this manner to facilitate indexing, retrieval, search, and analysis.

VI. SUMMARY

This paper introduced ontologies as an alternative type of data model to be compared and contrasted with the conventional, UML-based aviation data models that have been under development by government and industry over the past decade. Ontologies provide compelling new data modeling features and capabilities that apply to a wider set of use cases than conventional approaches. As technology and applications mature, we envision a viable path forward in which current standards are migrated and significantly enhanced through the eventual adoption of ontology-based data models.

REFERENCES

[1] G. Simsion and G. Witt, *Data modeling essentials*: Morgan Kaufmann, 2004.

[2] EUROCONTROL. *Flight Information Exchange Model*. Available: https://www.fixm.aero/

[3] J. Standley, V. Brown, P. Comitz, and J. Schoolfield, "SWIM segment 2 deployment and utilization in NextGen R&D programs," in *Integrated Communications, Navigation and Surveillance Conference (ICNS), 2012*, 2012, pp. G8-1-G8-5.

[4] B. K. Brunk and E. Porosnicu, "Aeronautical Information Exchange Model (AIXM) GIS interoperability through GML," in *25th Annual Esri International User Conference, San Diego, California, USA, 25th-29th July*, 2005.

[5] EUROCONTROL. *Aeronautical Information Exchange Model*. Available: https://www.aixm.aero/

[6] EUROCONTROL. *Weather Information Exchange Model*. Available: http://www.wxxm.aero/public/subsite_homepage/homepage.html

[7] IATA. *Airline Industry Data Model*. Available: http://www.iata.org/whatwedo/passenger/pages/industry-data-model.aspx

[8] D. Galgano and V. Simms, "An Intro to MMIXM," presented at the Air Transportation Information Exchange Conference, Silver Spring, MD, 2015

[9] S. Van der Stricht, "AMXM: Airport Mapping picking-up SWIM," presented at the Air Transportation Information Exchange Conference, Silver Spring, MD, 2015

[10] J. Rios, R. Jehlen, and Z. Zhu, "A spatial database for reroute planning," in *Digital Avionics Systems Conference (DASC), 2012 IEEE/AIAA 31st*, 2012, pp. 3E2-1-3E2-10.

[11] ICAO. (2013). *ICAO Meteorological Information Exchange Model (IWXXM)*. Available: http://www.icao.int/RO_APAC/Meetings/2013_MET_SG_17/IP15 - USA AI. 8.8 - MET exchange Model IWXXM.pdf

[12] EUROCONTROL. (2014). *FIXM Model Construction Guidelines Version 1.0 (1-Dec-2014)*. Available: https://www.fixm.aero/documents/ISS_CB0-2595241 - FIXM Model Construction Guidelines_1_0.pdf

[13] S. Wilson, R. Suzić, and S. Van der Stricht, "The SESAR ATM information reference model within the new ATM system," in 2014 Integrated Communications, Navigation and Surveillance Conference (ICNS) Conference Proceedings, 2014, pp. L3-1-L3-13.

[14] J. Rumbaugh, I. Jacobson, and G. Booch, *Unified Modeling Language Reference Manual, The*: Pearson Higher Education, 2004.

[15] T. R. Gruber, "Toward principles for the design of ontologies used for knowledge sharing," *International journal of human-computer studies,* vol. 43, pp. 907-928, 1995.

[16] B. Chandrasekaran, J. R. Josephson, and V. R. Benjamins, "What are ontologies, and why do we need them?," *IEEE Intelligent systems,* pp. 20-26, 1999.

[17] D. Fensel, *Ontologies*: Springer, 2001.

[18] T. R. Gruber, "The role of common ontology in achieving sharable, reusable knowledge bases," *KR,* vol. 91, pp. 601-602, 1991.

[19] T. Berners-Lee, J. Hendler, and O. Lassila, "The semantic web," *Scientific American,* vol. 284, pp. 28-37, 2001.

[20] W3C. (2014). *RDF 1.1 Concepts and Abstract Syntax*. Available: http://www.w3.org/TR/rdf11-concepts/

[21] W3C. (2014). *RDF Schema 1.1*. Available: http://www.w3.org/TR/rdf-schema/

[22] B. McBride, "The resource description framework (RDF) and its vocabulary description language RDFS," in *Handbook on ontologies*, ed: Springer, 2004, pp. 51-65.

[23] W3C. (2013). *SPARQL 1.1 Query Language*. Available: https://www.w3.org/TR/sparql11-query/

[24] W3C. (2014). *OWL 2 Web Ontology Language Document Overview (Second Edition)*. Available: https://www.w3.org/TR/owl2-overview/

[25] K. Rohloff, M. Dean, I. Emmons, D. Ryder, and J. Sumner, "An evaluation of triple-store technologies for large data stores," in *On the Move to Meaningful Internet Systems 2007: OTM 2007 Workshops*, 2007, pp. 1105-1114.

[26] W3C. (2014). *SPIN - Modeling Vocabulary*. Available: https://www.w3.org/Submission/2011/SUBM-spin-modeling-20110222/

[27] W3C. (2004). *SWRL: A Semantic Web Rule Language Combining OWL and RuleML*. Available: https://www.w3.org/Submission/SWRL/

[28] R. Fikes and A. Farquhar, "Distributed repositories of highly expressive reusable ontologies," *IEEE Intelligent systems,* pp. 73-79, 1999.

[29] J. Hartmann, R. Palma, and A. Gómez-Pérez, "Ontology repositories," in *Handbook on Ontologies*, ed: Springer, 2009, pp. 551-571.

[30] E. Gringinger, "Ontology-based representation and semantic querying of digital notices to airmen," in *Integrated Communications, Navigation and Surveillance Conference (ICNS), 2014*, 2014, pp. 1-23.

[31] M. Klauza, P. Czekalski, and K. Tokarz, "Air Traffic Data Integration using the Semantic Web Approach," 2014.

[32] B.-J. van Putten, S. R. Wolfe, and V. Dignum, "An Ontology for Traffic Flow Management," in *8th Aviation Technology, Integration, and Operations Conference*, Ankorage, AK, 2008.

[33] R. M. Keller, M. Kaplun, and L. Vizenor (eds.), "Semantic Web for Air Transportation Workshop Proceedings (SWAT-2015)," Washington, DC, 2015. Available: http://www.faa.gov/nextgen/programs/swim/governance/servicesemantics/#SWAT-Special-Interest-Group

[34] R. M. Keller, "NASA's ATM Ontology," presented at the Air Transportation Information Exchange Conference, Silver Spring, MD, 2015

[35] R. M. Keller, "Data Integration Using the NASA Air Traffic Management Ontology," presented at the Semantic Web for Air Transportation (SWAT-2015), Washington, DC, 2015. Available: https://www.faa.gov/nextgen/programs/swim/governance/servicesemantics/media/NASA ATM Ontology for SWAT revised.pdf

[36] R. M. Keller, S. Ranjan, M. Y. Wei, and M. M. Eshow, "Semantic representation and scale-up of integrated air traffic management data," in *International Workshop on Semantic Big Data*, San Francisco, California, 2016

[37] CAST/ICAO. *IACIS Aircraft Taxonomy*. Available: http://www.intlaviationstandards.org

[38] J. Cunningham and K. Hatton, "Traffic Management Initiative Ontology," presented at the Semantic Web for Air Transportation (SWAT-2015), Washington, DC, 2015. Available: https://www.faa.gov/nextgen/programs/swim/governance/servicesemantics/media/SWAT_TMIOntology_24Aug2015.pdf

[39] FAA. (2009). *Traffic Flow Management in the National Airspace System*. Available: https://www.fly.faa.gov/Products/Training/Traffic_Management_for_Pilots/TFM_in_the_NAS_Booklet_ca10.pdf

[40] D. Cowell, C. L. Buchanan, and J. Eberhardt, "Enterprise Information Management (EIM) at the FAA: Translating Semantic Technologies into Direct User Benefit," presented at the Semantic Web for Air Transportation (SWAT-2015), Washington, DC, 2015. Available: https://www.faa.gov/nextgen/programs/swim/governance/servicesemantics/media/EIM at the FAA SWAT 2015.pdf

[41] I. Niles and A. Pease, "Towards a standard upper ontology," in *Proceedings of the international conference on Formal Ontology in Information Systems-Volume 2001*, 2001, pp. 2-9.

[42] A. Gangemi, N. Guarino, C. Masolo, A. Oltramari, and L. Schneider, "Sweetening ontologies with DOLCE," in *Knowledge engineering and knowledge management: Ontologies and the semantic Web*, ed: Springer, 2002, pp. 166-181.

[43] V. Mascardi, V. Cordì, and P. Rosso, "A Comparison of Upper Ontologies," in *WOA*, 2007, pp. 55-64.

[44] A. Abdoullaev, *Reality, Universal Ontology and Knowledge Systems: Toward the Intelligent World: Toward the Intelligent World*: IGI Global, 2008.

[45] M. Kaplun, "FAA Web Service Description Ontological Model (WSDOM) - An Introduction," presented at the Semantic Web for Air Transportation (SWAT-2015), Washington, DC, 2015. Available:

https://www.faa.gov/nextgen/programs/swim/governance/servicesemant ics/media/FAA WSDOM Introduction.pdf

[46] FAA. *NAS Service Registry and Repository: Making Services Visible, Discoverable, and Understandable*. Available: https://nsrr.faa.gov/

[47] D. Martin, M. Paolucci, S. McIlraith, M. Burstein, D. McDermott, D. McGuinness*, et al.*, "Bringing semantics to web services: The OWL-S approach," in *Semantic Web Services and Web Process Composition*, ed: Springer, 2004, pp. 26-42.

[48] C. Uri, "FAA SWIM Controlled Vocabulary: A Semantic Approach to Clarifying SWIM Terminology," presented at the Semantic Web for Air Transportation (SWAT-2015), Washington, DC, 2015. Available: https://www.faa.gov/nextgen/programs/swim/governance/servicesemant ics/media/SWIM Controlled Vocabulary SWAT 8-24-2015.pdf

[49] A. Miles and J. R. Pérez-Agüera, "SKOS: Simple knowledge organisation for the web," *Cataloging & Classification Quarterly,* vol. 43, pp. 69-83, 2007.

[50] E. Gringinger, "SemNOTAM: Intelligent NOTAM Prioritization," presented at the Semantic Web for Air Transportation (SWAT-2015), Washington, DC, 2015. Available: https://www.faa.gov/nextgen /programs/swim/governance/servicesemantics/media/2015-08-24_semN OTAM_swat.pdf

[51] F. Burgstaller, D. Steiner, M. Schrefl, E. Gringinger, S. Wilson, and S. van der Stricht, "AIRM-based, fine-grained semantic filtering of notices to airmen," in *Integrated Communication, Navigation, and Surveillance Conference (ICNS), 2015*, 2015, pp. D3-1-D3-13.

[52] W. Zhu and L. Vizenor, "Cross-Organizational Semantic Services (CrOSS) and More," presented at the Semantic Web for Air Transportation (SWAT-2015), Washington, DC, 2015. Available: https://www.faa.gov/nextgen/programs/swim/governance/servicesemant ics/media/CrOSS Organization Semantic and More SWAT 2015.pdf

[53] A. Doan and A. Y. Halevy, "Semantic integration research in the database community: A brief survey," *AI Magazine,* vol. 26, p. 83, 2005.

[54] N. F. Noy, "Semantic integration: a survey of ontology-based approaches," *ACM Sigmod Record,* vol. 33, pp. 65-70, 2004.

[55] A. Borgida and P. F. Patel-Schneider, "A semantics and complete algorithm for subsumption in the CLASSIC description logic," *Journal of Artificial Intelligence Research,* pp. 277-308, 1994.

[56] C. Bizer, T. Heath, and T. Berners-Lee, "Linked data-the story so far," *Semantic Services, Interoperability and Web Applications: Emerging Concepts,* pp. 205-227, 2009.

[57] T. Heath and C. Bizer, "Linked data: Evolving the web into a global data space," *Synthesis lectures on the semantic web: theory and technology,* vol. 1, pp. 1-136, 2011.

[58] J. P. Holdren, "Expanding Public Access to the Results of Federally Funded Research," Office of Science and Technology Policy, February 22, 2013.

[59] E. Denney, "Towards an Ontological Basis for Aviation Safety Cases," presented at the Semantic Web for Air Transportation (SWAT-2015), Washington, DC, 2015. Available: https://www.faa.gov/nextgen /programs/swim/governance/servicesemantics/media/Safety Cases.pdf