Flight Information in a Post-Operations Big Data Environment

Vick G. Fisher
Introduction

- Introduce MITRE's Transportation Data Platform (TDP)
- Federal Aviation Administration’s (FAA) System-Wide Information Management (SWIM) System
- Storing SWIM Data in a Big Data Context
- SWIM Version Transition Planning and Execution
- SWIM Schema Evolution
Background: MITRE’s Transportation Data Platform

- Transportation is part of our *Nation’s Critical Infrastructure*, and a critical part of the FAA’s operations. Transportation is evolving quickly with autonomy, drones, artificial intelligence, and new business models injecting technology into operation.

- TDP is used to perform work required by MITRE Sponsors and Customers (FAA, Department of Homeland Security, Department of Defense, international civil aviation authorities, etc.)

- TDP focus: enabling increasing sophistication of data analyses, and reducing project time-to-value
TDP Operational Services

**Data Services**

- STDDS SMES
- NOP STARS
- SFDPS Flight
- TFM Flight
- SBS ADS-B
- ERAM Airspace
- NFDC Airport
- Airport Weather
- Airborne Weather

... 43 Data Services, 65 data products...

**User Services**

- TDP Core Algorithms
- Java API
- REST Web Service
- Clojure API
- Python API
- R API
- Jupyter
- R Studio
- Metadata Queries
- Key-Value Queries
- Data Documentation
- Service Desk

... ...

**Business Apps**

- AFS Analytics Dashboard
- IDEA Lab
- International
- ASIAS
System-Wide Information Management (SWIM) System Relation to FIXM

▪ “SWIM is the National Airspace System Data Sharing Backbone”
  – Replaces ad-hoc formats with standardized data formats
  – Replaces point-to-point connections with publish-subscribe model

▪ SWIM Services send data as XML
  – Some use Flight Information Exchange Model (FIXM)
  – Some use historical data formats
  – Some provide optional conversion to FIXM

▪ Implemented with Java Message Service (JMS)
Post-Operations Data: MITRE Archives Data For Current and Future Use

- Several years of raw, fused, and derived data archived
- Data used to support research
  - Safety, efficiency of the National Airspace System (NAS)
  - Studies can be short-term or long-term
  - Studies may cover a single airport, or the whole country
  - Simulations to evaluate proposed changes to the NAS
- Required by our contract in order to perform the research needed by our sponsors
  - Not for redistribution or sale
Avro is a data serialization system
- Built with Hadoop processing in mind
- Schema stored with data
- Some built-in support for schema changes

“Splittable” file formats
- Big data processing reads blocks of records starting in many places throughout the same file simultaneously

Parquet is a compressed, columnar file format
- Parquet files are “splittable”
- Supports Avro
- Tools provided for use with Java, Python, C++, and others
Ingestion into TDP begins with SWIM messages read from XML files

**JAXB (Java Architecture for XML Binding)** SWIM message objects are converted into Avro-compatible Java objects

Java objects stored in Parquet files on HDFS (Hadoop Distributed File System)
SWIM Schema Changes: Planning and Execution

- FAA SWIM provides a set of data services
  - Examples:
    - SWIM Terminal Data Distribution System (STDDS)
    - Traffic Flow Management (TFM)
    - Time-Based Flow Management (TBFM)
  - Each SWIM service has a separate schema
  - Each SWIM service changes independently of others
  - The following slides discuss the process of implementing a schema change from both producer and consumer points of view
SWIM Schema Changes: FAA Support

- **Announce date for a schema change**
  - Typically a year or more in advance
- **Changes communicated to subscribers via**
  - Industry forums
  - Webinars
  - Email lists
  - Help desk
  - Release notes
- **FAA upgrades producer software**
- **FAA provides sample data files in new format**
- **FAA publishes new format on the FAA test network**
SWIM Schema Changes: Consumer Preparation

- **SWIM client software must be revised to handle new schema**
  - Update SWIM client software to capture new schema
  - Test SWIM client on sample files
  - Establish network connection to topic publishing new schema on FAA test network
    - Test revised SWIM client on new topic data

- **Revise any local data models to store new schema**
  - Relational database tables
  - Java classes
  - Other
SWIM Schema Changes:
Consumer Preparation (Cont'd)

- **Update client to handle both old and new versions**
  - Allows new client software to be started before the new data version is released
  - When the release happens, client handles the new schema version seamlessly

- **Notify analysts, data scientists, software developers, etc., of upcoming changes**

- **Update applications that use the data**
  - Ingest processes, algorithms, dashboards, etc.
  - Could be the most labor-intensive step
SWIM Schema Changes: D-Day

- FAA updates version of *production* SWIM data service

- FAA sometimes provides both old and new versions of same data simultaneously, after release date
  - Supports subscribers not ready for new format

- Subscribers enable updated clients
Schema Evolution

- **Effects of schema changes: Operational versus Post-Operational**
  - If an organization is only using data operationally, old versions can be “forgotten”

- **Post-Operational Data Usage**
  - Multi-year archive of data implies various versions of data must coexist
  - Unified data model stores years of data from multiple releases

- **Serialization formats that include the schema can be helpful for managing schema evolution (e.g. Avro)**

- **Analyses and algorithms spanning releases should take advantage of all available data in each version**
### Schema Evolution Example: Field Removed

- STDDS R4 removed the "ground indicator" field from an Airport Surface Detection Equipment, Model X (ASDE-X) message.

<table>
<thead>
<tr>
<th>Date</th>
<th>Ground Indicator (removed in new release)</th>
</tr>
</thead>
<tbody>
<tr>
<td>03/29/19</td>
<td>1</td>
</tr>
<tr>
<td>03/30/19</td>
<td>2</td>
</tr>
<tr>
<td>03/31/19</td>
<td>0</td>
</tr>
<tr>
<td>04/01/19</td>
<td>?</td>
</tr>
<tr>
<td>04/02/19</td>
<td>?</td>
</tr>
<tr>
<td>04/03/19</td>
<td>?</td>
</tr>
</tbody>
</table>

New release starts here

What should these values be?
How does subscriber handle the change?
- The ground indicator field will be maintained in the MITRE data model
  - Older data in archive still contains this value
- Change SWIM client software to:
  - Accept new message version
  - Set field to null in new data objects of this type
- Change application software to handle null field value
Schema Evolution Example: Field(s) Added

- STDDS R4 added Vx, Vy, and V-vertical fields to the Terminal Automation Information Service (TAIS) message

<table>
<thead>
<tr>
<th>Date</th>
<th>Vx</th>
<th>Vy</th>
<th>V-vertical</th>
</tr>
</thead>
<tbody>
<tr>
<td>03/29/19</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>03/30/19</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>03/31/19</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>04/01/19</td>
<td>1.1</td>
<td>2.2</td>
<td>3.3</td>
</tr>
<tr>
<td>04/02/19</td>
<td>1.2</td>
<td>2.3</td>
<td>3.4</td>
</tr>
<tr>
<td>04/03/19</td>
<td>1.3</td>
<td>2.4</td>
<td>3.5</td>
</tr>
</tbody>
</table>

What should these values be?

New release starts here
How does subscriber handle the change?

- Add new fields to data model
- Add new fields to SWIM client software
- Archive to contain different schemas with different fields for same data type
- Change application software to take advantage of new information
Other Types of Schema Changes

- Actual examples from SWIM Traffic Flow Management (TFM) upgrade from R10 to R13
  - Increase length of existing field
  - Add XML type specifier to field
  - Add more enumeration values to an existing enumeration
  - Shift location of an element
  - Change element type
  - Change required field to optional
Conclusions

- MITRE converts from SWIM XML to Avro-compliant Java objects
- Big data processing requires “splittable” file format (such as Parquet) to enable parallel processing
- Planning for new releases with schema changes is complex and time-consuming
- Strategies needed for storing and analyzing multi-year archives covering multiple versions of the data
References

SWIM Home Page

ATIEC Conference
https://www.faa.gov/air_traffic/flight_info/aeronav/atiec/

FIXM Standard
https://www.fixm.aero/fixm_nas_extension_421.pl

TFM Data Service Message Details

TFM Data Service Conversion from R10 to R13
https://cdm.fly.faa.gov/?page_id=2559

STDDS R4 Release Notes
https://nsrr.faa.gov/sites/default/files/stdds-adp/STDDS%20R4.0%20Release%20Notes.pdf
References (Cont’d)

STDDS FIXM Conversion Service
https://www.faa.gov/air_traffic/technology/swim/stdds/media/FIXM_Mediated_STDDS_Data_Overview_v2.pdf

Avro
http://avro.apache.org/docs/1.9.0/
https://www.ibm.com/analytics/hadoop/avro
https://en.wikipedia.org/wiki/Apache_Avro
https://www.oreilly.com/ideas/the-problem-of-managing-schemas

Avro Schema Evolution
https://docs.confluent.io/current/schema-registry/avro.html

Avro Schema Resolution
http://avro.apache.org/docs/1.9.1/spec.html#Schema+Resolution

Parquet File Format
https://dzone.com/articles/understanding-how-parquet
## Acronym List

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADS-B</td>
<td>Automatic Dependent Surveillance - Broadcast</td>
</tr>
<tr>
<td>ASDE-X</td>
<td>Airport Surface Detection Equipment, Model X</td>
</tr>
<tr>
<td>ASIAS</td>
<td>Aviation Safety Information Analysis and Sharing</td>
</tr>
<tr>
<td>ERAM</td>
<td>En Route Automation Modernization</td>
</tr>
<tr>
<td>FAA</td>
<td>Federal Aviation Administration</td>
</tr>
<tr>
<td>FIXM</td>
<td>Flight Information Exchange Model</td>
</tr>
<tr>
<td>HDFS</td>
<td>Hadoop Distributed File System</td>
</tr>
<tr>
<td>HITL</td>
<td>Human-in-the-Loop</td>
</tr>
<tr>
<td>IDEA Lab</td>
<td>Integration Demonstration and Experimentation for Aeronautics</td>
</tr>
<tr>
<td>JAXB</td>
<td>Java Architecture for XML Binding</td>
</tr>
<tr>
<td>JMS</td>
<td>Java Message Service</td>
</tr>
<tr>
<td>NAS</td>
<td>National Airspace System</td>
</tr>
<tr>
<td>NFDC</td>
<td>National Flight Data Center</td>
</tr>
<tr>
<td>NOP</td>
<td>National Offload Program</td>
</tr>
<tr>
<td>SBS</td>
<td>Surveillance Broadcast System</td>
</tr>
<tr>
<td>SFDPS</td>
<td>SWIM Flight Data Publication Service</td>
</tr>
<tr>
<td>SME</td>
<td>Subject Matter Expert</td>
</tr>
<tr>
<td>SMES</td>
<td>Surface Movement Event Service</td>
</tr>
<tr>
<td>STARS</td>
<td>Standalone Terminal Automation Replacement System</td>
</tr>
<tr>
<td>STDDS</td>
<td>SWIM Terminal Data Distribution System</td>
</tr>
<tr>
<td>SWIM</td>
<td>System-Wide Information Management</td>
</tr>
<tr>
<td>TAIS</td>
<td>Terminal Automation Information Service</td>
</tr>
<tr>
<td>TBFM</td>
<td>Time-Based Flow Management</td>
</tr>
<tr>
<td>TDP</td>
<td>Transportation Data Platform</td>
</tr>
<tr>
<td>TFM</td>
<td>Traffic Flow Management</td>
</tr>
<tr>
<td>UAV</td>
<td>Unmanned Aerial Vehicle</td>
</tr>
<tr>
<td>XML</td>
<td>eXtensible Markup Language</td>
</tr>
</tbody>
</table>
MITRE’s mission-driven teams are dedicated to solving problems for a safer world. Through our federally funded R&D centers and public-private partnerships, we work across government to tackle challenges to the safety, stability, and well-being of our nation.

Learn more [www.mitre.org](http://www.mitre.org)
NOTICE

This work was produced for the U.S. Government under Contract DTFAWA-10-C-00080 and is subject to Federal Aviation Administration Acquisition Management System Clause 3.5-13, Rights In Data-General, Alt. III and Alt. IV (Oct. 1996).

The contents of this document reflect the views of the author and The MITRE Corporation and do not necessarily reflect the views of the Federal Aviation Administration (FAA) or the Department of Transportation (DOT). Neither the FAA nor the DOT makes any warranty or guarantee, expressed or implied, concerning the content or accuracy of these views.

© 2019 The MITRE Corporation. All Rights Reserved.