Flight Information in a Post-Operations Big Data Environment

Vick G. Fisher

MITRE

© 2019 The MITRE Corporation. Approved for Public Release; Distribution Unlimited. Public Release Case Number 19-2875

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

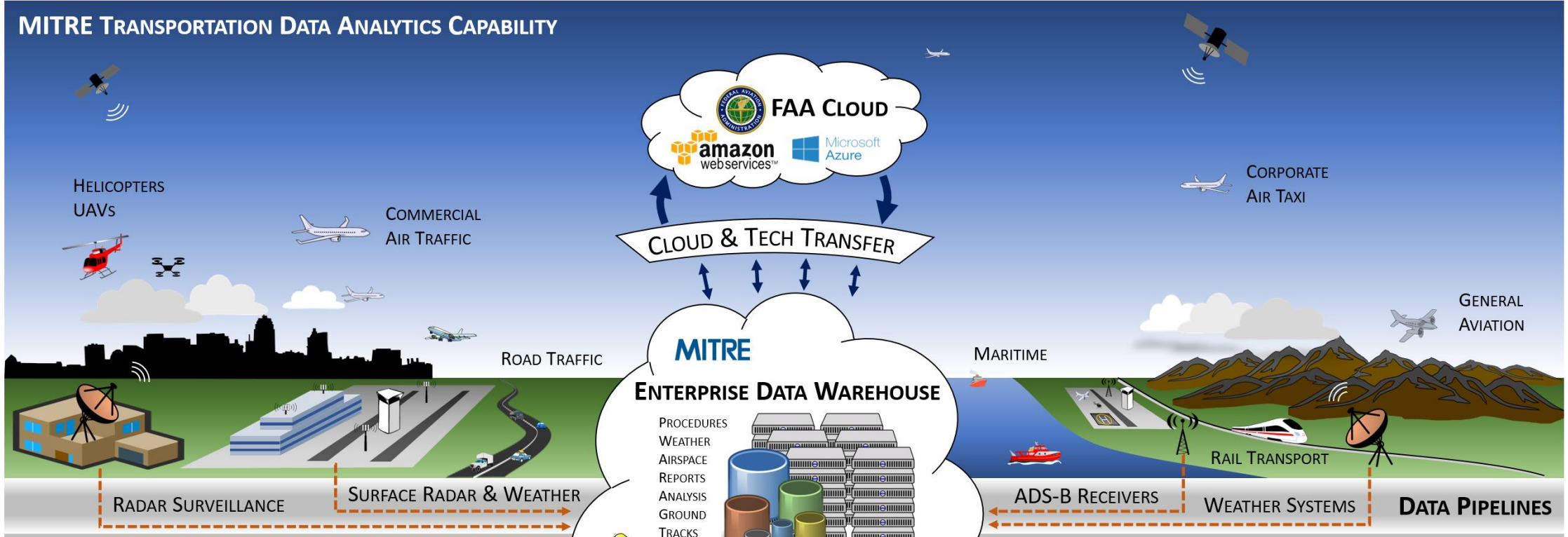
© 2019 The MITRE Corporation. All rights reserved.

MITRE

Background: MITRE's Transportation Data Platform

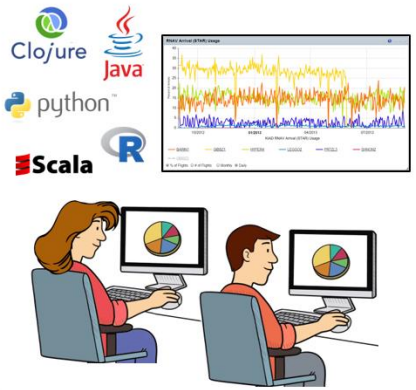
- Transportation is part of our ***Nation's Critical Infrastructure***, and a critical part of the FAA's operations. Transportation is evolving quickly with autonomy, drones, artificial intelligence, and new business models injecting technology into operation.
- TDP is used to perform work required by 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

MITRE Transportation Data Analytics Capability



MITRE Analysts

- Research – Development
- Analysis - Algorithms
- Visualizations - Trending

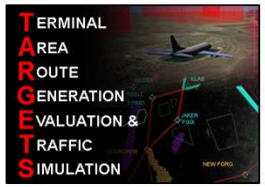


Scalable Data Processing

Data Access APIs & Services

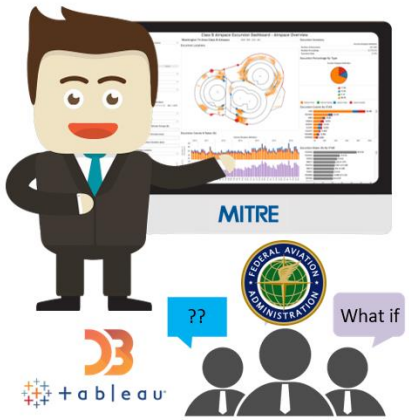
MITRE Projects

- Simulations - Design
- HITLS - Optimization
- Modeling - Assessments



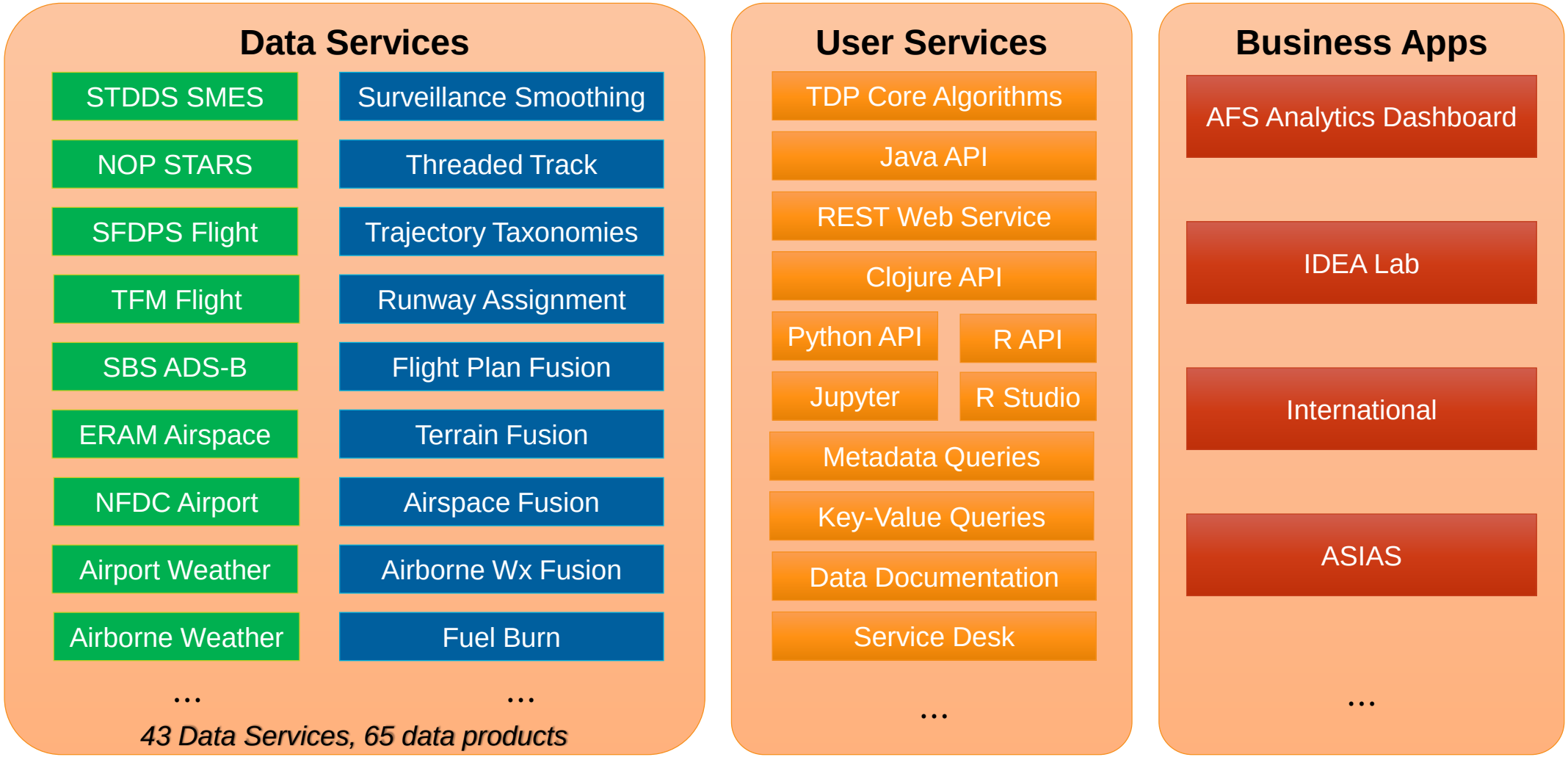
Applied Analytics

- Sponsor Requests
- Key Performance Indicators
- Dashboards



MITRE

TDP Operational Services



MITRE

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)**

© 2019 The MITRE Corporation. All rights reserved.

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

© 2019 The MITRE Corporation. All rights reserved.

MITRE

TDP Storage Technologies and Concepts

- **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

© 2019 The MITRE Corporation. All rights reserved.

TDP Storage Format

- **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)

© 2019 The MITRE Corporation. All rights reserved.

MITRE

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

© 2019 The MITRE Corporation. All rights reserved.

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**

© 2019 The MITRE Corporation. All rights reserved.

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

© 2019 The MITRE Corporation. All rights reserved.

MITRE

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

© 2019 The MITRE Corporation. All rights reserved.

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**

© 2019 The MITRE Corporation. All rights reserved.

MITRE

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**

© 2019 The MITRE Corporation. All rights reserved.

Schema Evolution Example: Field Removed

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

Date	Ground Indicator (removed in new release)
03/29/19	1
03/30/19	2
03/31/19	0
04/01/19	?
04/02/19	?
04/03/19	?

← New release starts here

← What should these values be?

© 2019 The MITRE Corporation. All rights reserved.

MITRE

Schema Evolution Example: Field Removed (Cont'd)

- **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

© 2019 The MITRE Corporation. All rights reserved.

Schema Evolution Example: Field(s) Added

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

	Fields Added in New Release		
Date	Vx	Vy	V-vertical
03/29/19	?	?	?
03/30/19	?	?	?
03/31/19	?	?	?
04/01/19	1.1	2.2	3.3
04/02/19	1.2	2.3	3.4
04/03/19	1.3	2.4	3.5

← What should these values be?

← New release starts here

© 2019 The MITRE Corporation. All rights reserved.

MITRE

Schema Evolution Example: Fields Added (Cont'd)

- **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

© 2019 The MITRE Corporation. All rights reserved.

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

© 2019 The MITRE Corporation. All rights reserved.

MITRE

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

© 2019 The MITRE Corporation. All rights reserved.

References

SWIM Home Page
https://en.wikipedia.org/wiki/System_Wide_Information_Management
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
https://cdm.fly.faa.gov/?page_id=1991
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

© 2019 The MITRE Corporation. All rights reserved.

MITRE

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

© 2019 The MITRE Corporation. All rights reserved.

Acronym List

Acronym	Definition
ADS-B	Automatic Dependent Surveillance - Broadcast
ASDE-X	Airport Surface Detection Equipment, Model X
ASIAS	Aviation Safety Information Analysis and Sharing
ERAM	En Route Automation Modernization
FAA	Federal Aviation Administration
FIXM	Flight Information Exchange Model
HDFS	Hadoop Distributed File System
HITL	Human-in-the-Loop
IDEA Lab	Integration Demonstration and Experimentation for Aeronautics
JAXB	Java Architecture for XML Binding
JMS	Java Message Service
NAS	National Airspace System
NFDC	National Flight Data Center
NOP	National Offload Program
SBS	Surveillance Broadcast System
SFDPS	SWIM Flight Data Publication Service
SME	Subject Matter Expert
SMES	Surface Movement Event Service
STARS	Standalone Terminal Automation Replacement System
STDDS	SWIM Terminal Data Distribution System
SWIM	System-Wide Information Management
TAIS	Terminal Automation Information Service
TBFM	Time-Based Flow Management
TDP	Transportation Data Platform
TFM	Traffic Flow Management
UAV	Unmanned Aerial Vehicle
XML	eXtensible Markup Language

© 2019 The MITRE Corporation. All rights reserved.

MITRE

MITRE

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

© 2019 The MITRE Corporation. All rights reserved.

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.

© 2019 The MITRE Corporation. All rights reserved.