

Department of the Air Force

Integrity - Service - Excellence

Integrating Data Science & Cost Analysis



May 2022



Bad News

- **Generally speaking, the cost community is way behind the curve from a data science perspective vs. current state-of-the-art**
 - Our **PROCESSES** are outdated
 - Often, each analyst/team of analysts separately pulls down data from various sources which is repetitive & monotonous
 - Lack of “flattened” structure/format for data, lack of consistent methods/analysis
 - Result is inconsistent, compartmentalized and unstructured datasets and “ad-hoc”/stove-pipped analysis
 - Depending on the organization, modern **TOOLSETS** are either non-existent or under-utilized, and inconsistent across different organizations
 - **PRESENTATIONS** to leadership are often static vs. dynamic



Good News

■ We have more data than ever before

- For DoD – Complete overhaul of CSDRs with “FlexFiles”
 - Cost data
 - Software data
 - Technical data
 - Programmatic data
- Advana +750 data sources
- Sites like DACIMS, PMRT, EVM-CR etc.

■ As a community (generally speaking) we have the necessary building blocks to succeed in the field of data science

- Mathematicians, engineers, as well as a variety of other technical backgrounds
- Many of our job responsibilities already overlap with those in the data science field

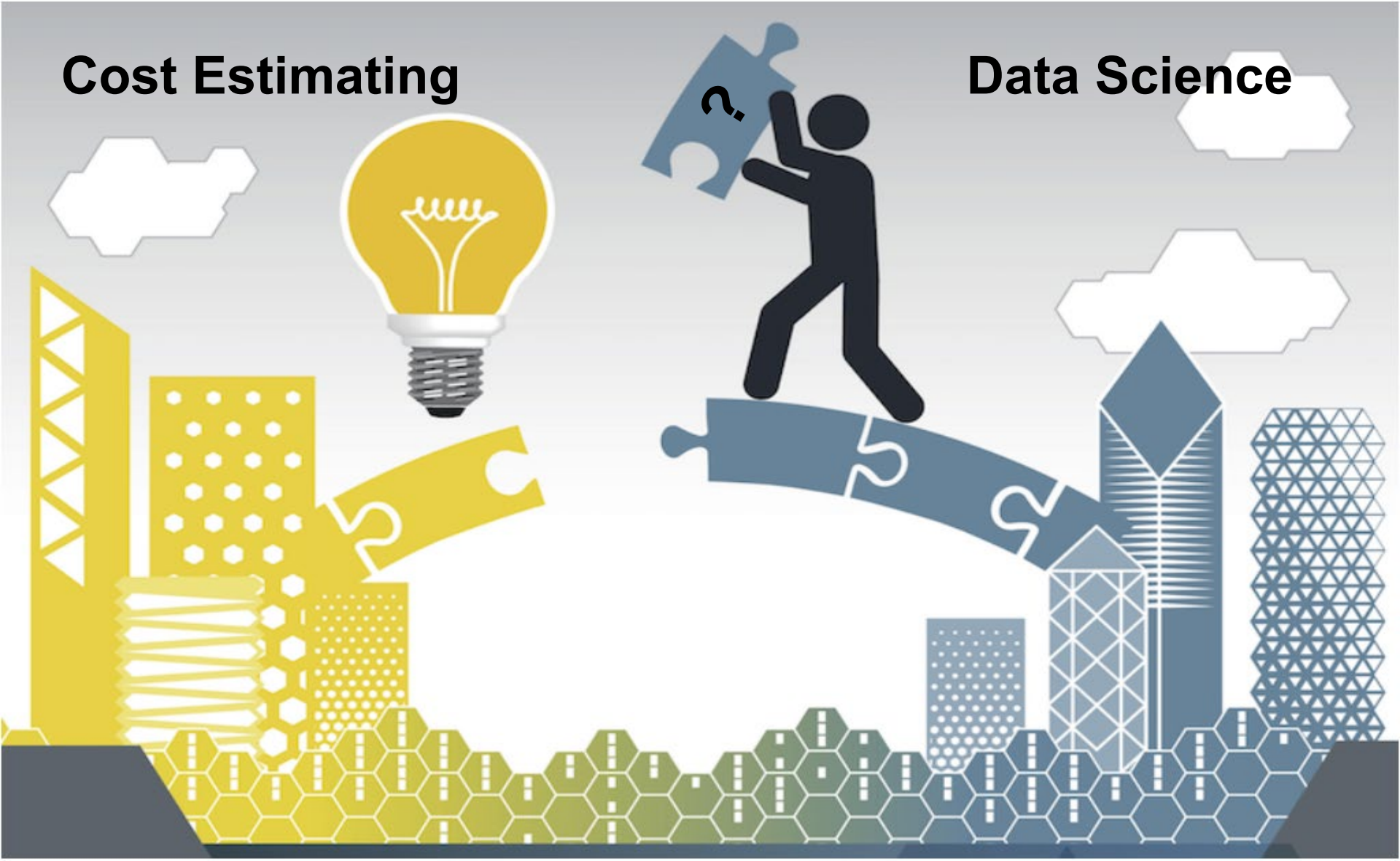


Bridging the Gap

Cost Estimating



Data Science





VAULT

■ What is the VAULT?

- A cloud-based platform designed to make data more Visible, Understandable, Linked and Trusted (VAULT)
- Run by the Chief Data Office of the Air Force but available to anyone with a CAC on a government network
- The real value of the Vault (and a MUST if proprietary data is involved) is the ability to establish access-controlled “tenant” instances/spaces
- Provides analysts with very powerful software tools that help us to centralize, automate and modernize our processes and models



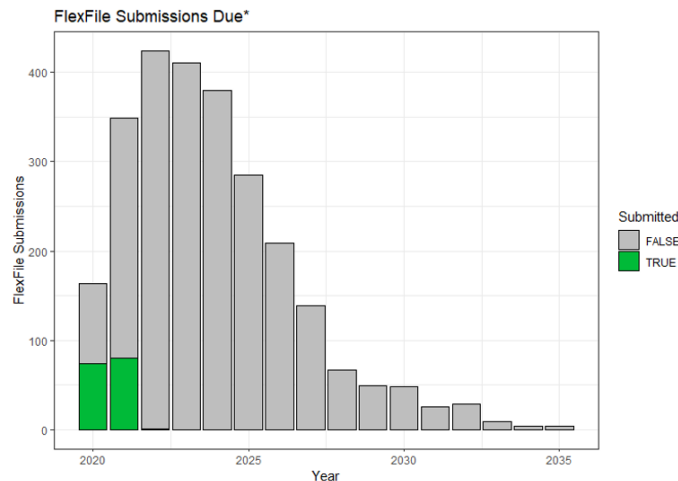
VAULT, cont.

- **Identified the VAULT, an Air Force platform, as the optimal solution for AFCAA at this time**
 - Proprietary data protection – only AFCAA personnel have access to our tenant space
 - VAULT obtained Authority to Operate (ATO) for IL5 in 2019 and for SIPR in 2020
 - Emphasis on analytic tool accessibility and allowance for “Bring Your Own Tools” – a significant requirement/benefit to cost community
 - AFCAA custom models/applications in VAULT cannot currently be replicated in another platform without capability degradation
 - AF Operations Research community platform of choice – Enhancing AF cost community synergistic Analytics Sharing
 - Future development priorities are aligned with cost community needs



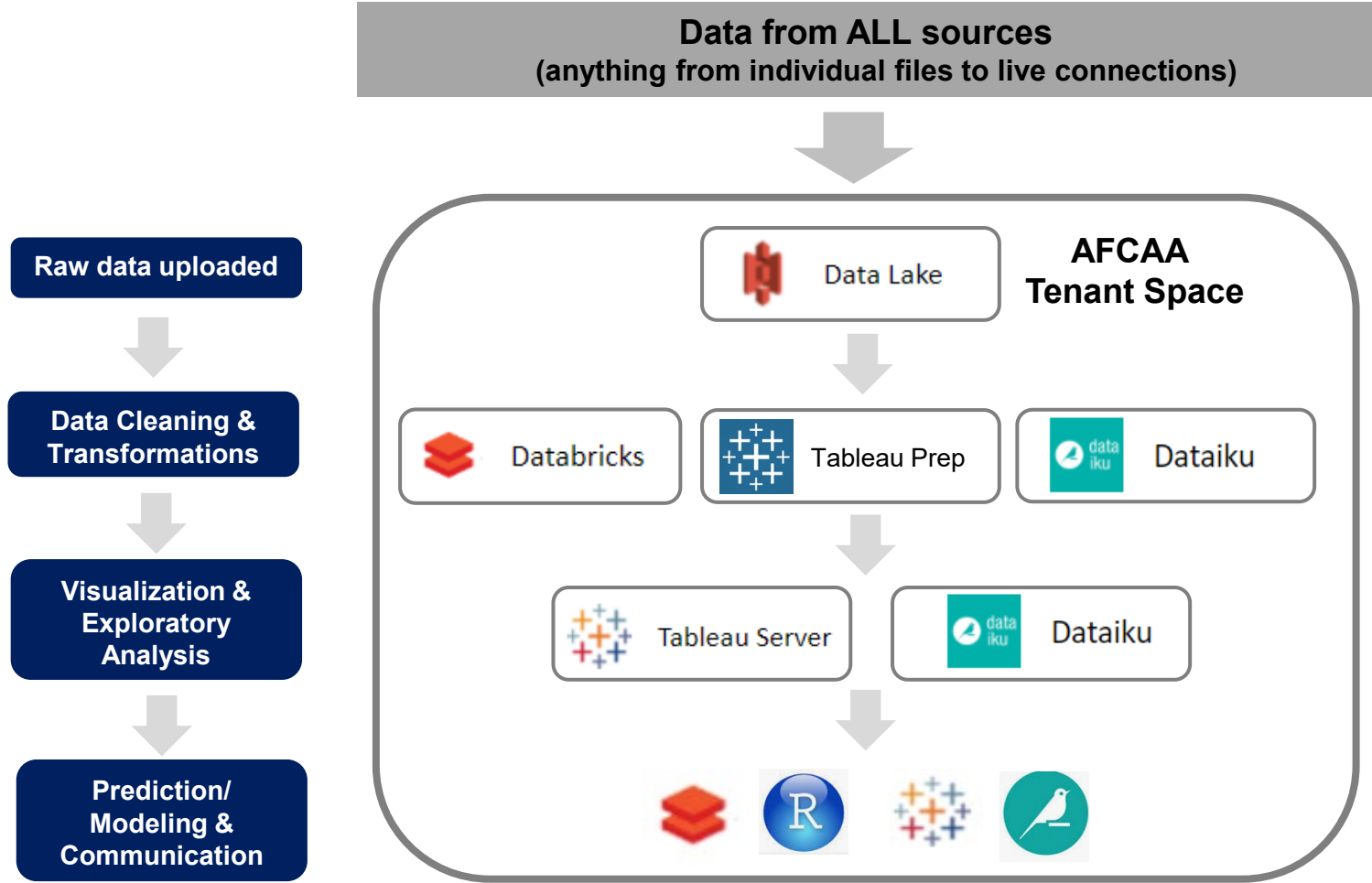
Value of VAULT

- Given the large number of FlexFiles expected over the next few years, we need to adjust to be ready to accommodate
- VAULT has the capability to help our analysts “end-to-end” from ingesting data all the way to performing analysis with advanced tools
 - Complete traceability from raw data to actual analysis
 - Opportunity for automation, collaboration and centralization
 - Eliminates version issues with tools





How AFCAA is Using the VAULT



Turn manual, stove-piped processes into automated, centralized and efficient workflows



Realized Benefits

Legacy

Traceability is dependent on documentation & process used by analyst

Only saved versions are kept – can lose trace to data in certain versions of models if not properly handled

Often have issues with compatibility of desktop versions

Mostly manual steps – not easily repeatable and often not well documented

Extremely difficult to get desktop tools approved on high side

Performance limited to desktop compute

Models are tedious to update and are often several years outdated

Org-wide changes like inflation updates have to be individually updated in each model manually

Power of data science in cost community often limited due to very few programmers in the field

VAULT

Complete step by step traceability to original, raw data

Insight into who made change & when – and can revert back to a previous version

No compatibility issues with different versions of desktop software once in the cloud

Automated steps from raw data to final product so that it's repeatable on new data that's received

Can replicate environment on the high side (SIPR now, JWICS 2022)

Performance will scale based on compute available in "cluster"- equivalent of groups of machines

Can update models much more easily

Have the ability to centralize updates so that analysts can pull them in to models quickly and easily

Data science tools in a no code/low code environment



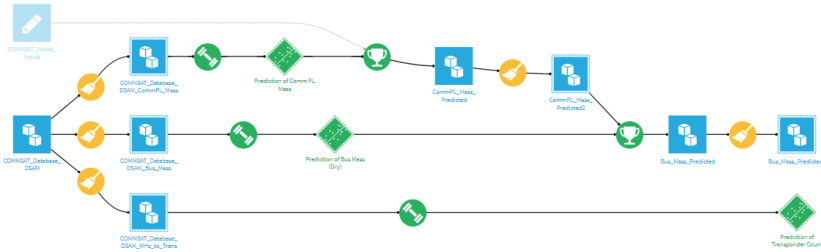
Summary of Data Analytic Activities

- **AFCAA currently Testing/Piloting in VAULT**
 - Centralization/automation of activities that are typically “stove-piped” and manual such as:
 - Data cleansing and analysis
 - Custom-built data visualizations/dashboards with raw data from various sources consolidated in one location
 - Development and use of cost models connected to technical/schedule models
 - Development and use of custom built cost applications
 - Ability to alleviate issues with tool access and processing power
- **Internal Governance Processes and Rules of Engagement have been drafted and continue to be revised based on testing/pilot findings**
- **Live Demo of activities in the VAULT**



Live Demo

Live Demo of AFCAA activities in the VAULT



Program Overview Dashboard

Program:

- Program Overview
- Technical Data
- EVM Data
- Budget Data
- Contract Data
- CSDR Data
- SDSR Data

Advanced Extremely High Frequency Satellite (AEHF) is a joint service satellite communications system that provides global, survivable, secure, protected, and jam-resistant communications for high priority military ground, sea and air assets. The system consists of four operational satellites in Geosynchronous Earth Orbit that provide 10 times the capacity of the 1990s-era Military Strategic and Tactical Relay Block II satellites. The system provides continuous 24-hour Extremely High Frequency Extended Data Rate coverage between 65 degrees north and 65 degrees south latitude. AEHF allows the National Security Council and Combatant Commanders to control their tactical and strategic forces at all levels of conflict up to and including general nuclear war, and it supports the attainment of information superiority.

The AEHF operational system is composed of three segments: space, terminals, and mission control. The space segment consists of a cross-linked constellation of satellites to provide worldwide coverage. The terminal segment includes fixed and mobile ground terminals, ship and submarine terminals, and airborne terminals. The mission control segment controls satellites on orbit, monitors satellite health, and provides communication system planning and monitoring. This segment is also survivable, with both fixed and mobile control stations.

Timeline

Category	1999	2000	2001	2002	2003	2005	2007	2008	2009	2012	2013	2014	2015	2016	2017	2018	2019	2020
Program																		
Space																		
Ground																		

Interactive Scoring

ADD TO COMPARATOR COMPARE (0) ...

Filter...

Design Life (yrs)

Satellite Class (HBC, MU, Mil, NASA-Exp, HTS, Dual)

Total Bandwidth All Transponders (MHz)

Primary Band

Program Type/Sponsor

Prediction for Total Comm Payload Mass (lbs):

Most influential features for Total Comm Payload Mass (lbs) (ICE)

Primary Band	High Influence
Design Life (yrs)	Low Influence
Total Bandwidth All Transponders (MHz)	Low Influence
Satellite Class (HBC, MU, Mil, NASA-Exp, HTS, Dual)	Low Influence
Program Type/Sponsor	Low Influence

Contract Data Dashboard

Most Category

- ADMINISTRATIVE
- BASISLINE
- CSDF
- SCHEDULE
- TECHNICAL
- UNKNOWN



Backup



Tools Matrix (Draft)

- Identifies some (not all) tools and how they can benefit us in each stage of the cost estimating cycle

Primary Function	Tools	Available in Vault	Available on AFCAA Desktop	ETL	Visualize/ Exploratory Data Analysis	Model/ Regression Analysis	Model/ Machine Learning	Document/ Communicate	User Friendliness/ Ease of use*	*Languages Supported/ Required
Primarily Statistical	CO\$STAT		Y							
	JMP		Y							
Primarily Visual	PowerBI		Y			*	*			R/ Python supported
	Tableau Pro+Prep	Y	Y			*	*			R/ Python supported
	Qlik					*	*			
Notebook based	Rstudio	Y (in Databricks)	Y							R required
	Jupyter	Y (in Dataiku)								Python required
End-to-end	Dataiku	Y								R/ Python/SQL supported
	R Shiny	Y								R required
	Plotly Dash	Y								Python required
	Databricks	Y								R/ Python/SQL required

*Not inherent to tool but supported with incorporation of scripts/modules

Capability	User-friendliness
Good	Easy
OK	Medium
None	Requires programming



Data Analytics User Group Overview

■ What this User Group **IS**

- Analysts from a wide spectrum of different government organizations that are **CURRENT users of advanced data analytic tools** and can represent their organization by talking specifically about them and ideally be able to demonstrate exactly how they are using those tools
- Liaising with Data Tools Tiger Team (whose mission is to better inform leadership on what is needed to identify, procure and adopt the right tools for the cost community)
- Government civilians
- Contractors directly supporting a government organization
- Meets regularly every 3 weeks

■ What this User Group **IS NOT**

- Providing training for novice users
- Making authoritative decisions about which tools should be used
- Industry contractors



Data Analytics User Group Goals

■ Short term Goals :

- Create a community of analysts using data analytics tools to collaborate
- Discuss each tool in detail to include the different ways that each organization is using the tools to their advantage
- **Demonstrate and share** results with the group so we can consolidate best practices and lessons learned
- Collaborate in order to **avoid duplicative data analytic efforts** and leverage work that has already been done to the greatest extent possible
- More widespread outreach to the cost community to help with **adoption of tools**

■ Potential longer-term goals:

- Collaboration on products to be eventually hosted in the DTM Hub

Email me for more information or to request to join: sarah.green.10@us.af.mil