

NATIONAL RECONNAISSANCE OFFICE

CDMS: Developing a Solution for Cost Data Management

ICEAA 2019 Forum

May 2019



SUPRA ET ULTRA



Agenda

- Problem Statement
- CDMS Overview
- CDMS Development Plan
 - Database
 - Tagging
 - Website
- Team Composition/Strategy
- Future Development
- Questions?



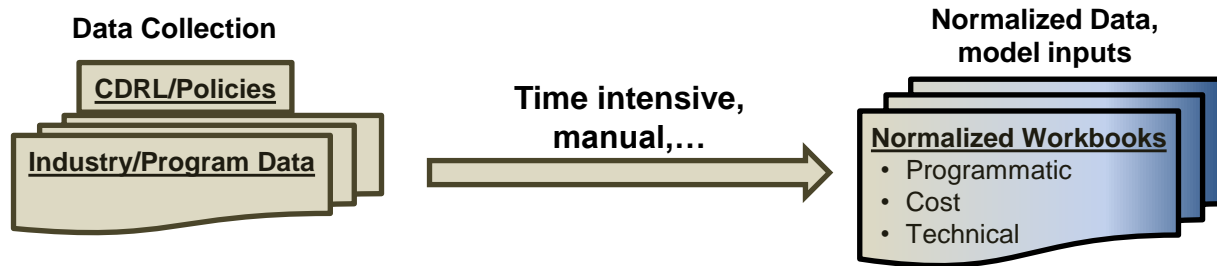
NRO CAAG Data Management System (CDMS)

- A strong data infrastructure is essential to enable current and future analytics for cost estimating organizations
 - NRO is developing the CAAG Data Management System (CDMS)
 - Data migration is underway with IOC planned for this Summer
- Existing data, models, tools, user requirements, and operating environment are important considerations when selecting a development approach for a Data Management System
- A hybrid agile development has proven effective
 - Steering Group of management and cost estimators define requirements and ensure CDMS is meeting needs
 - Weekly meetings with Steering Group and Developers facilitates communication, guides development, and resolves issues
 - Complex calculations and functionality concerns are quickly addressed
- Data tags and a robust user interface enable situational awareness, efficiencies, and advanced analytics
 - Key to unlocking data science and advanced analytics



The Starting Point (Problem Statement)

- NRO CAAG has 300+ datasets on programs spanning 50+ years
 - Covers entire enterprise: space, ground, communications, etc.
 - Data collected from contractors in native softcopy formats



- CAAG normalizes data by program in spreadsheets for use in cost estimating and methods development
- No existing repository that connects normalized data
- No continuous linkage of data to methods

We started from square 1!

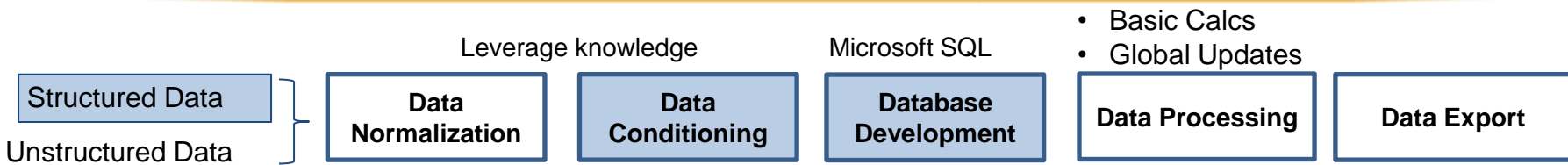


Why Develop a NRO CAAG Data Management System?

- CAAG normalized data is presently managed in spreadsheets and stored in a shared folder structure
 - Neither of these solutions are scalable, maintainable or adequately flexible to meet current analyst needs
 - Spreadsheets are hard to query and are not meant to be used for data storage
 - Better technologies exist to accomplish this
- A Data Management System (DMS) is the solution
 - Defined by the requirements delivered in this task
 - More than a data repository...a DMS can perform consistent normalization procedures and store that data in a more practical format
- A DMS needs to be scalable, maintainable, flexible and traceable
 - The range of formats and total amount of CAAG data will continue to increase (e.g. EVM Center of Excellence, Program Assessment, Enterprise Analysis)



Structured Data Approaches

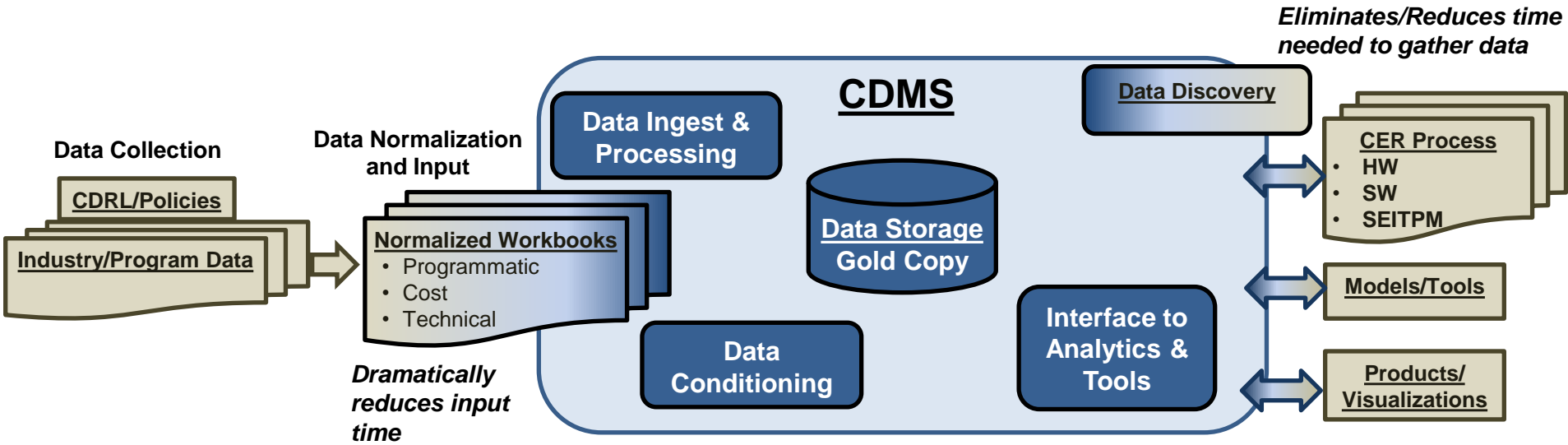


• **Requirements express need for highly customized solution connecting NRO normalized data with NRO derived CERs with NRO custom models**

Approach	Advantage	Disadvantage
Status Quo	<ul style="list-style-type: none"> No additional effort 	<ul style="list-style-type: none"> Does not meet requirements Leaves inefficient processes in place
Flexfiles/CADE	<ul style="list-style-type: none"> Standard across DoD High level reports, maps data to 1921 reports (DoD standard) 	<ul style="list-style-type: none"> More library like, vice NRO requirements Data not mapped to NRO CERs, models & Tools Little to no control over development
COMM/CSO	<ul style="list-style-type: none"> Focus on grouping data and visualizations <i>Processes unstructured data (Potential future CAAG unstructured component)</i> 	<ul style="list-style-type: none"> Less capability to solve CAAG data associations to CERs, Models, & Tools Not as effective for current priority needs
ODNI Data Infrastructure	<ul style="list-style-type: none"> Provides associations between data and end items Less time to stand-up 	<ul style="list-style-type: none"> ACCESS may not handle CAAG needs (2GB limit) Not capable of access control (security/sharing) Still requires customization for CAAG processes & tools Less control or coordinated development
Custom SQL Relational DB	<ul style="list-style-type: none"> Built to align to NRO tailored requirements Full control and easiest approach to make modifications Scalable solution 	<ul style="list-style-type: none"> Time to develop database solution



CAAG CDMS Environment



- Extensive Data Set**
- NRO Space programs
 - NRO Ground programs
 - Other Space programs

- Database Features**
- Programmatic, cost, and technical data
 - SEITPM, Subsystem, and box HW data
 - Defines data tags to enhance analytics
 - Implements automated normalization
 - Incorporates macro change capability
 - Inflation, SWBS
 - Establishes a web based environment

- Future Capabilities**
- Enhance GUI
 - Expand CDMS to incorporate additional data sets
 - Integrate CAAG models and tools
 - Advance cross-cutting analytics and data mining capabilities

CDMS is a key infrastructure element driving operations and processes to advance CAAG capabilities



What is CDMS?

- The CAAG Data Management System is the CAAG's solution for aligning its data, methods, tools and the Standard Work Breakdown Structure (SWBS) into a single database
- The NRO CAAG started development of CDMS in the Summer of 2017
 - Included realignment our complete Normalized Workbook (NWB) data history into a standard format before synchronizing it with CDMS
 - Included NRO, other government agencies, and commercial programs
- Applies a SWBS with cost and technical data down to the box level (level 5+)
- A standardized process where data is normalized, tagged, and validated to establish the official cost and technical resource for each program

CDMS will contain normalized data for NRO programs and enhance analytics



CDMS Development Overview (Phases)

Work to Date

Future Expansion

Capability Layer (Requirements, Definition, Architecture, Documentation, and Management)

- Define system requirements
- Define data needs and data dictionary
- Develop CDMS template to import historical data
- Define data tags to enhance analytics
- Determine reporting requirements
- Develop CDMS Manual with user guide
- Perform V&V, and Testing
- Train users

- Coordinate CDMS migration to cloud
- Identify requirements for CAAG estimating data sets
 - New CAAG business areas and the associated data sets
 - Unstructured data sets (i.e. CAAG library)
- Define interface requirements for CAAG Models and Tools
 - NRO CAAG custom, other NRO, and commercial products
- Explore data science, analytical, and visualization capabilities
 - Leverage data science, machine learning

Key

Capability Plan

Development Plan

Data Plan

Development Layer (Database, Coding, Interfaces, Etc.)

- Establish Database Environment
- Build Database for:
 - Programmatic, cost, and technical data
 - Space and Ground data
 - SEITPM, Subsystem, box HW
 - NRO, DoD, Commercial, and other programs
 - SEITPM and Subsystem data
- Implement automated Normalization
- Incorporate macro change capabilities
- Establish web based environment

- Migrate CDMS to cloud environment
- Enhance GUI to support all CAAG business areas
- Build out CDMS for additional CAAG estimating data sets
 - New CAAG business areas and the associated data sets
 - Unstructured data sets (i.e. CAAG library)
- Build analytical interfaces
 - Custom NRO CAAG models and tools
 - Other NRO databases (financial, EVM, etc.)
 - Commercial analysis tools
 - Cloud data science and advanced analytical tools

Data Layer (Add Legacy NRO, USCM, and Commercial Program Normalized Workbook data)

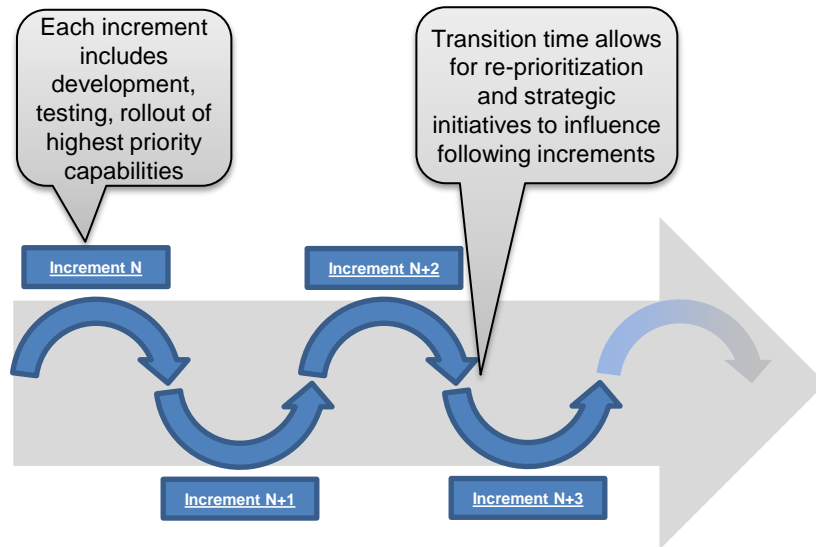
- Import data:
 - NRO Space programs
 - NRO Ground programs
 - Non-NRO programs

- Import new CAAG data sets
 - 25+ data sets currently in models and tool



CDMS Development Strategy and Team Composition

Hybrid Agile Approach



- CDMS is being developed in an incremental fashion
 - Quarterly with requirements review/prioritization
- Incremental development allows for:
 - Prioritization updates
 - Achievable goals
 - Focused development
- Steering group has been successful in overseeing development:
 - Drive testing requirements and QC
 - Solicit suggestions from CAAG analysts
 - Provide feedback on development prioritization
 - Develop future capability requirements

Steering Group

- Brian Wells (CDMS Lead)
- 5-6 cost estimators

Technical

- 1-2 cost estimators

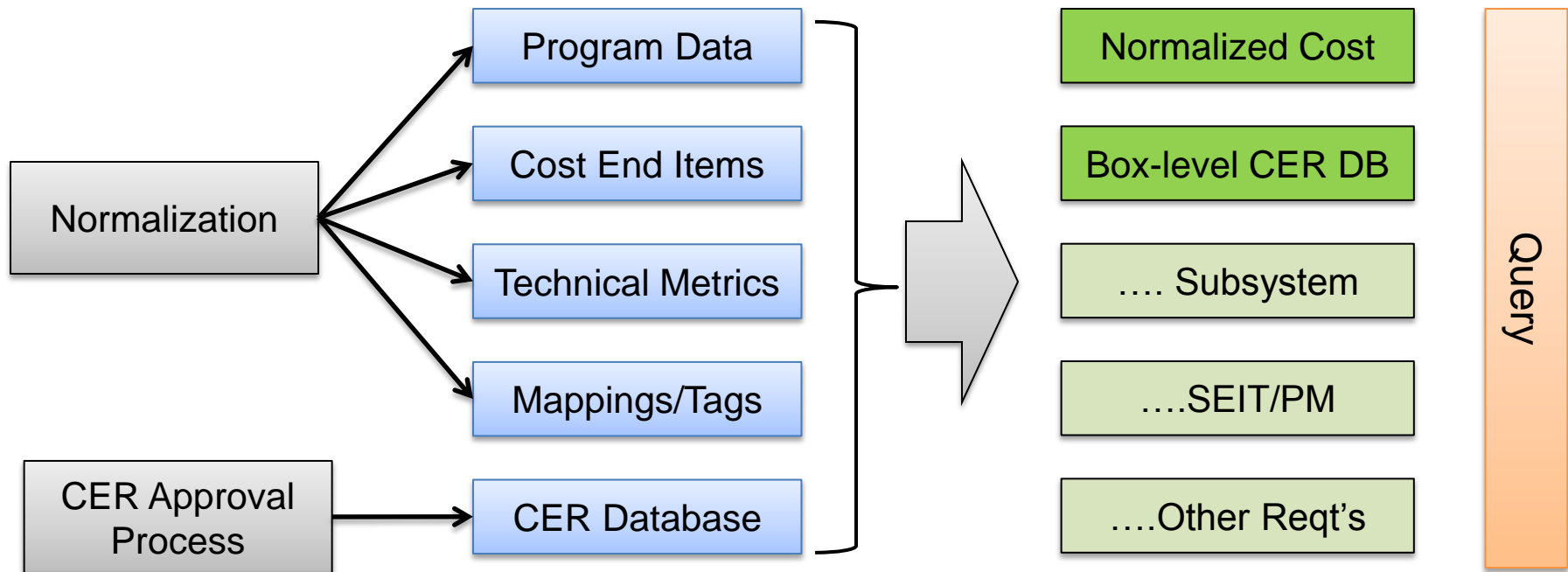
SW Development

- 3-4 developers

Cost & Acquisition Assessment Group (CAAG)



Flow of Data into Proof of Concept



- Focus on core capabilities: Data conditioning and Database Development
- Identify limitations and benefits which only are seen at detailed implementation level

Proposed data structure leverages existing products and proven processes



Tagging Schema

Program Data

Contract and Technical Summary	Unit	Input	Comments and References
Contract Variables			
Program Name		XXX	XXX
Contract Scope		XXX	XXX
Contracting Agency or Directorate		XXX	XXX
Contract Number		XXX	XXX
Contractor		XXX	XXX
Contract Type		XXX	XXX
Contract Value without fee	STYM	XXX	XXX
Contract Value with fee and COM	STYM	XXX	XXX
G&A base		XXX	XXX
CMY-NR	year	XXX	XXX
CMY-R	year	XXX	XXX
CMY-Total	year	XXX	XXX
Percent subcontracted	%	XXX	XXX
Contract Authority to Proceed	date-mon-yr	XXX	XXX
Contract CDR	date-mon-yr	XXX	XXX
Contract End Date	date-mon-yr	XXX	XXX
Period of Performance	mos	XXX	XXX
Percent complete	%	XXX	XXX

Cost End Items

End-Items	Hours				\$FY00K			
	NR	NR	NR	NR	R	R	R	R
	Direct Eng Hours	Direct Mfg Hours	Direct Labor Hours	Cost Thru G&A	Direct Eng Hours	Direct Mfg Hours	Direct Labor Hours	Cost Thru G&A
BATT	10	10	10	2,000	10	10	10	2,000
BATT CTL ASSY	20	20	20	3,000	20	20	20	3,000
BOOSTER ADAPTER-A FRAME	30	30	30	4,000	30	30	30	4,000

Tagging/Mapping

Program Specific End Item	Program Specific SWBS Mapping	Program Specific Box-Level Unique ID	Standard Hardware Type
BATT	1.2.2.4.11	Rechargeable Batteries	Battery - NIH2
BATT CTL ASSY	1.2.2.4.12	Charge Control Electronics	Junction Box (EPS)
BOOSTER ADAPTER-A FRAME	1.2.5	Booster Adapter	Booster Adapter

Technical Metrics

Box-Level Unique ID	Unit Wt	% WG	NSY	Weight per Sys	E	TQ	PC	F	S	R	%Uniq Mech	%Uniq Elec	%New Mech	%New Elec	HW Type	% Sub	Compo site %UD	Compo site %ND	DEV Units	Rec Units
Rechargeable Batteries	10.0	0	1	10	0.0	0.0	0.0	1.0	0.5	0.0	20.0	0.0	0.0	0.0	EM	100	16.0	0.0	-	1.5
Charge Control Electronics	20.0	0	1	20	0.0	0.0	0.0	1.0	0.5	0.0	95.0	17.0	0.0	0.0	E	0	32.6	0.0	0.2	1.5
Booster Adapter	30.0	0	1	30	0.0	1.0	0.0	1.0	0.5	0.0	80.0	0.0	70.0	0.0	M	0	80.0	70.0	1.0	1.5

All data is notional for presentation purposes

CER Database

Unique ID	REC CER Name	REC % Error	NR CER Name	NR % Error	Hardware Type	CER Date	NR CER Status	REC CER Status	Dependent Variable in '00	CER Function
Program X Rechargeable Batteries	REC_Battery05NIH		NR_BatteriesNIH213	-72%	Battery - NIH2	2013	Approved	Approved	NR(FY00SK) =	$\alpha * (T1(BY00SK)/1000^a) * (BPC^b) * (\%ND^c) * (T1(BY00SK)/1000^d) * Harm$
Program X Charge Control Electronics	REC_EPSElectronics05		NR_Misc_Elex13	-4%	Junction Box (EPS)	2013	Approved	Approved	NR(FY00SK) =	$\alpha * (T1(BY00SK)/1000^a) * (BPC^b) * (\%ND^c) * (T1(BY00SK)/1000^d) * Harm$
Program X Booster Adapter	REC_BoosterAdapter06		NR_Structure13	-99%	Booster Adapter	2013	Draft	Draft	NR(FY00SK) =	$\alpha * (T1(BY00SK)/1000^a) * (BPC^b) * (\%ND^c) * d^e * structure * e^f * Harm + Da$



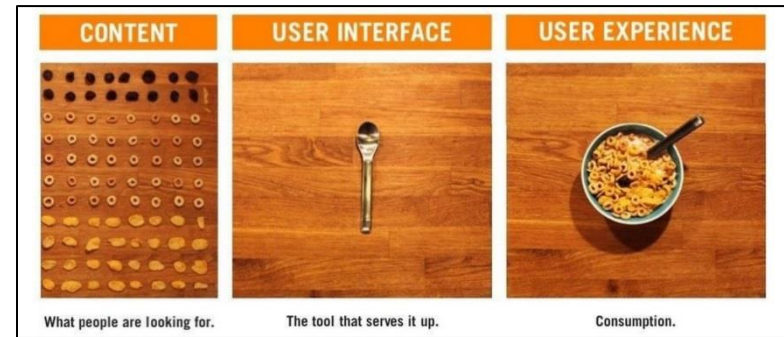
Why Tag?

- Tagging involves relating data to one another via pre-determined types and lists
 - Manual and initially time intensive
- SWBS is a key “tag”, but there are a lot of useful ways to slice and dice such a large dataset
 - I.e. Equipment type, Application, Orbit, Contractor
- Tagging allows/improves:
 - More standardized model development
 - Efficient queries on detailed datasets (i.e. CER development)
 - Better selection of analogy programs for cost estimates
 - Improved data awareness



GUI Development

- User interface is just as important as functionality
- Key benefit of database solution is integration to the website
 - Allows for easier access and editing of data
 - Up/download of data from verified sources
 - Data dashboard (future development)
 - Repository Integration (i.e. SharePoint)
- Facilitate easier sharing of data
 - Defined security controls to protect classified and proprietary data
 - Physical distance sharing
 - Common interface points



Credit: <https://www.bwired.com.au/blogs/the-importance-of-user-interface-and-why-it-is-critical-to-your-success-online>

Example of Everyday GUI



<https://mutcd.fhwa.dot.gov>

“If you build it wrong, they will come...and leave”



CDMS Website Screenshots



BENJAMIN TRUSKIN'S PROFILE PROGRAMS OUTPUT ADMINISTRATION HELP LOGOUT

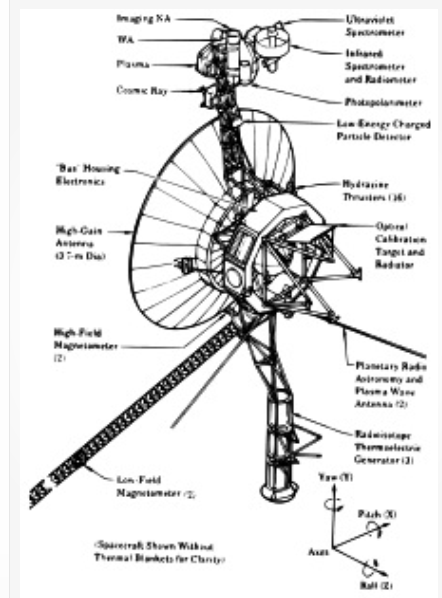
TAG BANK		Manage Tag Bank
List Name		Users
Aligned Setlpm Level		Manage Programs
Aligned Setlpm Type		Manage Inflation
Aligned Subsystem Hardware Type		Update Tag Bank
Aligned Subsystem Tech Driver		Update Tag Bank
Aligned Subsystem Type		Update Joining Tag Bank
Contractor		Update Tag Bank
Aligned Contractor		Update Joining Tag Bank
Hardware Type		Update Tag Bank
Hardware Type Tech Driver		Update Joining Tag Bank
Setlpm Level		Update Tag Bank
Setlpm System Tech Driver		Update Tag Bank
Setlpm System Type		Update Tag Bank
Setlpm System Type Tech Driver		Update Joining Tag Bank
Setlpm Type		Update Tag Bank
Subsystem Tech Driver		Update Tag Bank
Subsystem Type		Update Tag Bank
Tech Driver		Update Tag Bank

Voyager (ICEAA Example)

SUMMARY UPLOAD DATA STATUS CHANGE LOG VERSIONS ISSUES 0

Description:
The Voyager spacecraft each weigh 773 kilograms (1,704 pounds). Of this total weight, each spacecraft carries 105 kilograms (231 pounds) of scientific instruments...

Contract Name:
Contract Number:
DPM: No DPM
Backup DPM: Benjamin Truskin
NWB Data Status: No Data



BOX-LEVEL REPORT FILTER

Hardware Type:

Subsystem:

Standard WBS Number:

Show Child WBS Number Elements

Tailored SWBS Description:

Prime Contractor:

Normalized By: CAAG

Unit Weight:

%ND:

%UD:

NR (BY00\$K):

REC (BY00\$K):

Total (BY00\$K):

Select Programs to Filter By:
*If no program is selected all programs will be searched

Select Contracting Agency/Organization

Select Contracting Agency/Directorate



Select All Unselect All



Future Development/Other Improvements

- Data Situational Awareness
 - Identification of data sparseness, types, sources, quality
 - Dashboards to manage data collection and resolve data issues
- Mass Database updates
 - Inflation
 - Normalization methodology
 - SWBS
- Incorporate additional CAAG data sets
- Integration/Interface with CAAG models and tools
- Cloud migration
- Enable data science and advanced analytics
 - Integration with commercial data analysis tools



Questions?

Contact Information

- Government POC: Brian Wells
 - 571-304-8890, wellsdb@nro.mil
- Tony Occhiuzzo, Integrity Applications Incorporated
 - 571-304-8873, occhiuan@nro.mil
- Benjamin Truskin, Technomics
 - 571-304-8858, truskinb@nro.mil

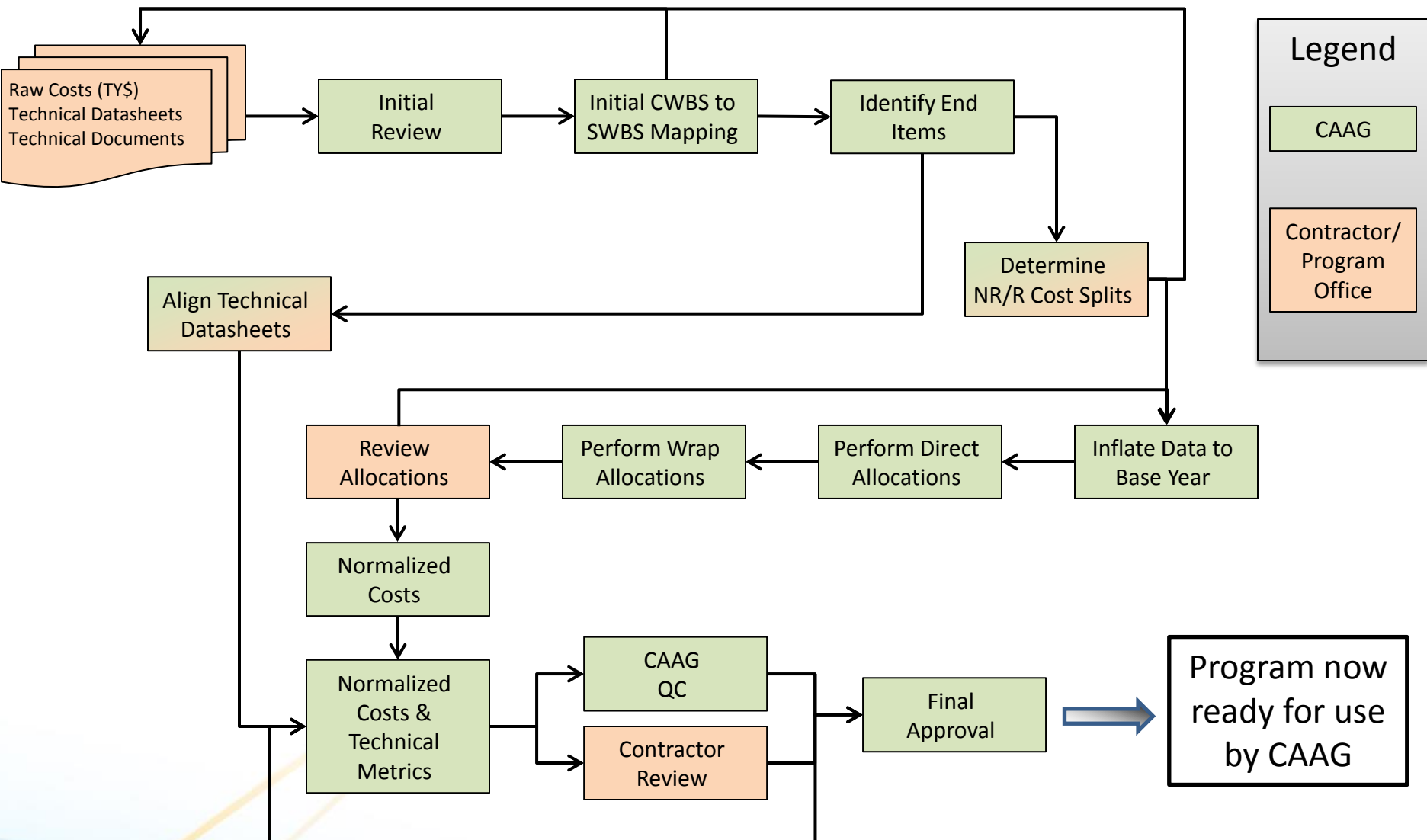
NATIONAL RECONNAISSANCE OFFICE

SUPRA ET ULTRA





CAAG Normalization Process (Simplified)





CDMS Technology Overview

- Database Engine: MS SQL Server
- Web/Application Server: Apache, Tomcat
- Languages: Java, Groovy, VBA, SQL
- Frameworks: Grails, POI, Spring
- IDEs: IntelliJ, MS SQL Server Management Studio
- Configuration Management: S2P Bitbucket on SourceTree
- Collaboration & Issue Tracking: S2P Confluence & JIRA