



Applied automated data reduction tools:

Developing automated data reduction tools provide consistent and repeatable data analysis and reduced tester/analyst workloads while increasing results confidence

Angela Roe and Robert Posey, TASC, Inc.

▶ Agenda

- Problem Space
 - Overview
 - Distributed Common Ground/Surface System (DCGS)
 - Defense Readiness Reporting System-Strategic (DRRS-S)
 - Coalition Interoperability Assurance & Validation (CIAV)
- Solution Approach
- Solution Description
 - Rule-Based Processing
- Customer Benefits
- Summary

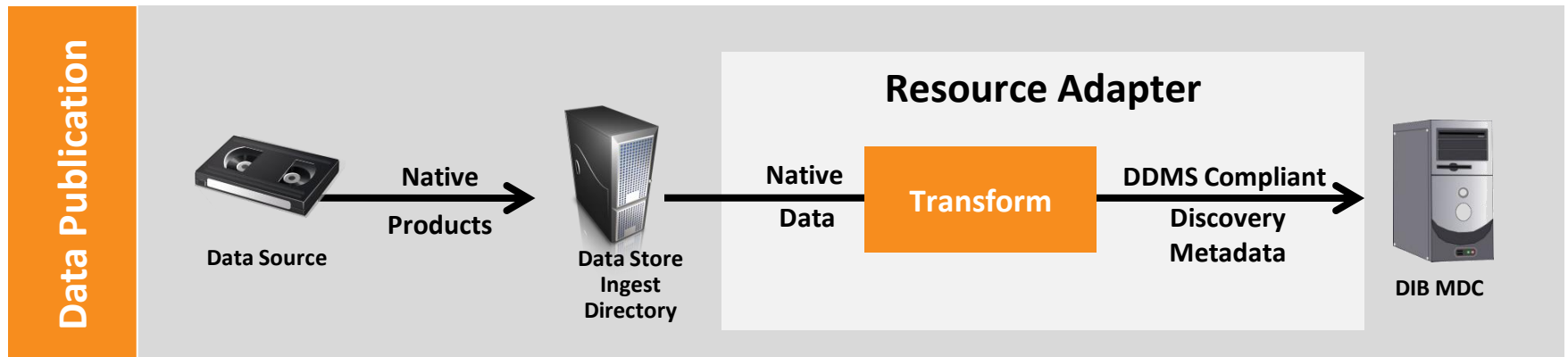
► Problem Space - Overview

- Significant impacts to Systems Under Test (SUT)/Families of Systems (FOS) issue identification efforts
 - Limited testing time
 - Increased system complexity
 - Increased overall testing costs
 - Increased data sample size requirement
 - New/updated Protocols
 - Analyzing data conversions from one to many different protocols/formats (Transformation Analysis)

These impacts significantly reduced by effectively applying advanced testing technologies, such as rule-driven automated data reduction capabilities.

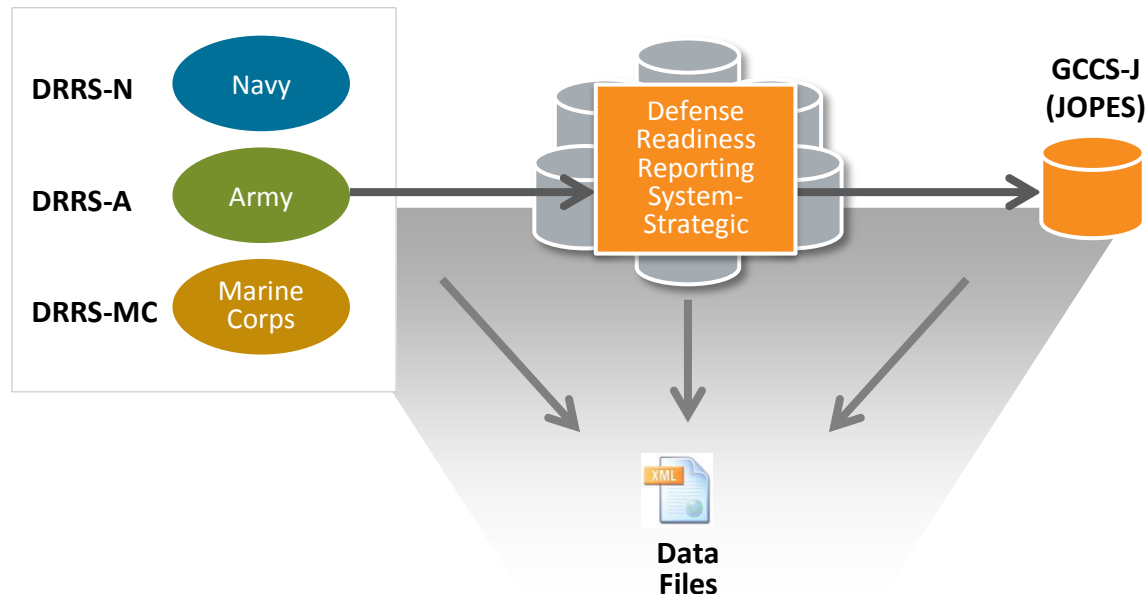
► Problem Space - DCGS

- Required assessing metadata accuracy, consistency, and completeness through transformation analysis
- Previous data publication and analysis process
 - Manually extracted files from data store and DCGS Integration Backbone (DIB) Metadata Catalog (MDC)
 - Individually validated source and Discovery Metadata (DMD) files (Compliant Image Validation Analyzer (CIVA), XML Spy)
 - Manually correlated file pairs to verify crosswalk



► Problem Space – DRRS-S

- Required comparisons of transformed information as data propagated through systems
 - Verify data transformations occur correctly between systems
 - Simultaneously compare and report on multiple data files
 - Increase the data sample size that could be analyzed
 - Multiple data file collection points



▶ Problem Space - CIAV

- Required Instrumentation and Analysis during CIAV Events
 - Change from analysis of screen captures to 100% of data collected and available for analysis
 - Comparisons of transformed information as data propagated through systems
 - Verify data transformations occur correctly between systems
 - Increase the data sample size that could be analyzed

► Solution Approach

- Apply rule-driven automated data reduction capability
 - Accelerated testing and data analysis processes
 - Allowed for incremental testing during the SUT/FOS lifecycle
 - Reduced testing risks by increasing the amount of initial testing, regression testing, and analysis completed
 - Expanded test sample sizes in some cases by 200 to 400 percent
 - Improved repeatability of results by standardizing analysis efforts and reporting

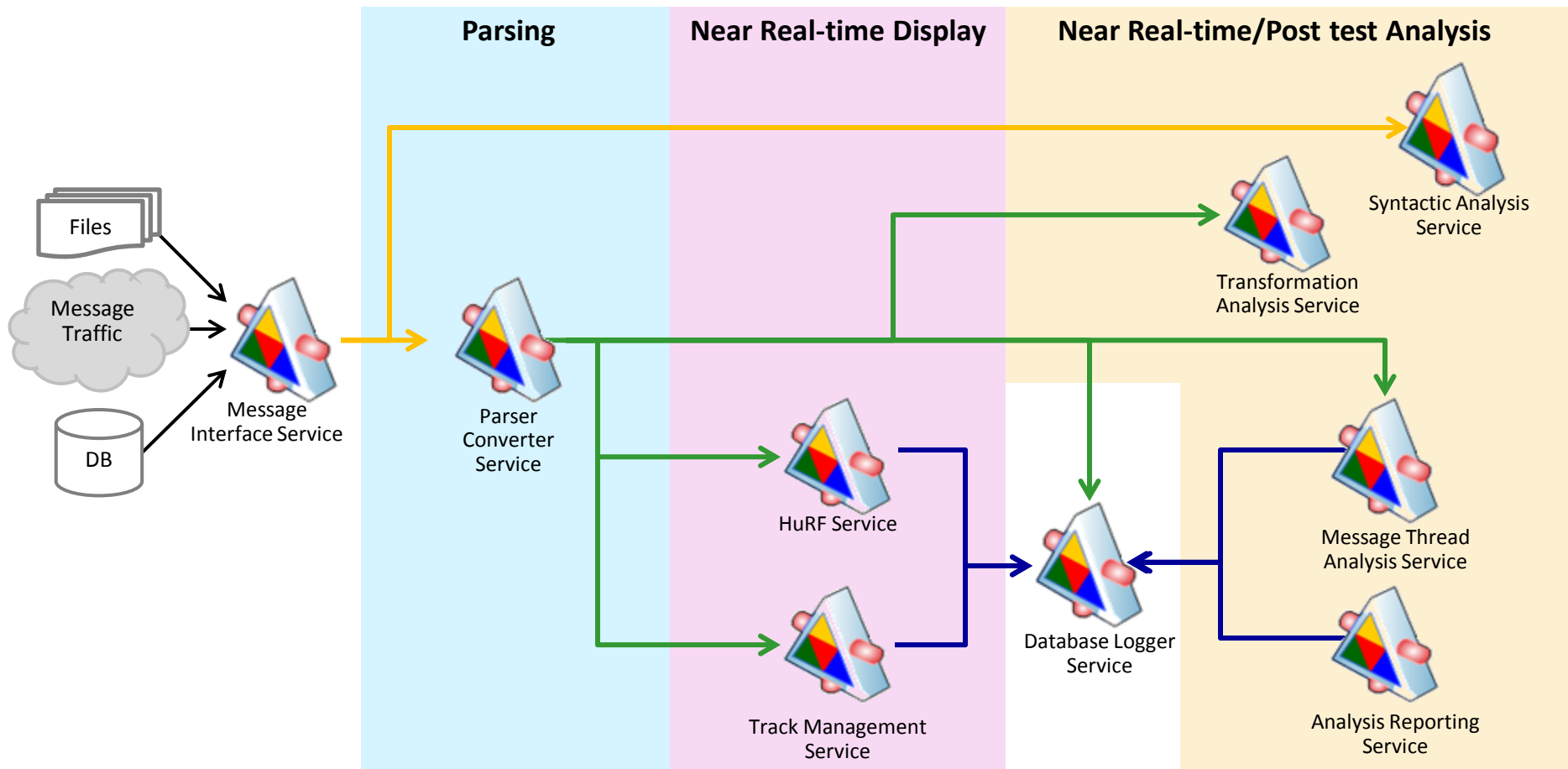
- ✓ Automated and Rules-driven
- ✓ Increased the degree of confidence in the results
- ✓ Lowered the cost of test conduction and analysis
- ✓ Significantly reduced risks in testing

► Solution Description

- Developed non-protocol specific processing, analysis, and display engines in the Joint Analysis Net-centric Evaluation Testing Toolkit (JANETT)
 - Used extensible markup language (XML)-based “rule” files stored as Standard Query Language (SQL) Database (DB) Files to process, analyze, and display received data
 - Modifications to supported protocols required minimal to no software development support to create, maintain, and execute rules
 - Developed a rules repository
 - Efficiently maintain, analyze, and modify the rules

- ✓ Reduce Overall Cost while increasing testing capabilities
- ✓ Allowed quicker completion of the Information Assurance (IA) review process
- ✓ Rule file updates completed by non-programmers reducing costs
- ✓ Cross-project rules utilization

▶ Rule-Based Processing



- Process multiple protocols at once

- Change rules on-the-fly to adapt to data structure changes
- Quickly identify data version changes

- Quickly compare multiple 'like' or 'unlike' files
- Provide consistent assessment results

▶ Customer Benefits

- Quickly addressed data file changes
 - Updated DRRS-S rule files ready within 15 minutes of identifying input data file changes
 - Data reprocessed and ready for analyst review in as little as an hour
- Allowed for quickly comparing multiple ‘like’ files
 - Reduced the average time to compare DCGS file pairs to less than 30 minutes per pair
 - previously averaged ~ 2-4 hrs/file pair

▶ Customer Benefits

- Allowed for quickly comparing multiple 'unlike' files
 - Improved flexibility in comparing data fields between files containing different data types, precision, and sizes
 - Developed 15 rules from scratch in less than 4 hours which covered processing, analyzing/comparing, and reporting on ~200 data messages from 2 different databases
- Quickly identify possible differences in data conversions as data moved through systems (transformation analysis)
 - Customers notified of provided data file or transmitted message changes within minutes of receiving the data

▶ Customer Benefits

- Provided consistent assessment results
 - Mitigate human error
 - Reduced inconsistent crosswalk interpretation
- Ability to automate comparisons between multiple sources of data simultaneously
 - Improved flexibility in assessing multiple data types
- Increased ability to assess more over time
 - Allowed for conducting more analysis using larger sample sizes and test files

▶ Customer Benefits

- Reduce IA review process time when updates occur
 - Data files perceived differently than software
 - Data file updates after installation normally only incur nominal IA review
 - Software updates take 2-3 weeks or longer to move through the IA process
- Reduce cost to support new/different versions
 - Rule creation/updates do not require software engineers
 - Rules can be shared between projects reducing costs
 - Verification/Validation/Accreditation different between software and rules
 - Rule changes may only need to be validated

▶ Summary

- Reduced
 - Tester and analyst workloads
 - Dependencies on analyst skill set
 - Possibility of human error during analysis
 - Costs to support new/updated protocols

- Increased
 - Data sample size that could be analyzed
 - Degree of confidence in the results
 - Analysis consistency

- Allowed
 - Quicker completion of the IA review process
 - Cross-project rules utilization

▶ POCs

- Ms. Angela Roe, TASC, Inc.
– Angela.Roe@TASC.com
- Mr. Robert Posey, TASC, Inc.
– Robert.Posey@TASC.com