GOVERNMENT BETA
The Value of Unscripted Testing

Jonathan Haulund
Chief Software Engineer
United States Air Force

Rocky Khullar
Project Engineer
The Aerospace Corporation

Greg Dawson
Associate Vice President
LinQuest Corporation
Overview

- Introduction
- System Architecture
- Approach and Challenges
- Unscripted Test
- Day-in-the-Life Test
- Incremental Test
- Risk Management: Quality vs. Progress
- The Contract and Award Fee Plan
- Final Thoughts
Introduction

“Government Beta”
The Beta Effect

If you want to see your software do what it was designed to do, put it in the hands of its users (literally and figuratively)
The Process of Development

- Requirements
- Design
- Implementation
- Verification
- Deployment
- Maintenance

Cost of Change
How do we avoid “beta” releases?

The Incremental Effect

Government Acceptance

Government Acceptance

Government Acceptance

Government Acceptance

Cost of Change

Increment n

Increment n+1

Increment n+2

Increment n+3
The Government’s Role

- Develop Requirements
- Develop Request for Proposal
- Award Contract
- Oversee Execution
  - Manage Risk
  - Manage Contract Modifications
  - Manage Award Fee/Incentives
- Accept Product
- Oversee Sustainment

*Best Opportunities to Influence Outcome
The Pursuit of Quality/Suitability

Cost of Change

Requirements
Design
Implementation
Verification
Deployment
Maintenance

Completeness, Accuracy & Language
Prototyping
Design/Reuse Assessment
Design to Test
Unit Test
Unscripted Test
Confidence Demo
“Day in the Life” Test
Government Test
Extensible Test Framework

Improvements not made
Improvements made
Government
System Architecture
Our System – Advanced EHF

Legend

Communication With Legacy System
Communication With New System

Mission Control Segment (MCS)

Legacy Satellite (Milstar)
New Satellite (AEHF)
New Satellite (AEHF)

Backward Compatible Crosslink
Enhanced Capacity Crosslink

Test and Training Simulation Element
Operations Support & Sustainment Element
Mission Operations Element
Mission Planning Element

Legacy Terminal
New Terminal

User Terminal Segment

Legend

New
Legacy System

Communication With Legacy System
Communication With New System
Mission Planning Element (MPE)

- Responsible for planning, scheduling, execution, data, and monitoring satellite and terminal resources
- **Multiple users (100’s) greatly varying in knowledge, skills and mission**
- Extensive reuse of software from a previous system
- High complexity due to payload/terminal designs
  - 100’s of screens, interdependent data
  - > 1.5M SLOC in C/C++
  - Software is classified SECRET
MPE High Level Conops

- Hundreds of MPSS
- Distributed Planning
  - From National Command Authorities, through all Echelons, to Field Users
- One MPE Application Set
- Coordination Through Data Distribution

Many Users / Many Different Missions

SOC/SOM
- Plan Strategic Networks
- Develop UCC Apportionments
- Distribute Apportionment to UCCs
- Process Resource Utilization Data
- Analyze Resource Utilization Data

UCC/RSSC
- Sub-Apportion Resources
- Distribute to UCCs
- Process Resource Utilization Data
- Analyze Resource Utilization Data

Navy Component

Air Force Component

Marine Component

Army Component

Army Corps

Fence/Comm/ Antenna Controllers/ Terminals
- Re-point Antennas
- Establish Fences
- Activate Nets
- Develop Detailed Plan
- Generate Terminal DB and Execution Plan Data
- Analyze Resource Utilization Data

Antenna Pointing/
Fence
Commands/
Activate Nets

GOVERNMENT BETA – The Value of Unscripted Testing
MPE Mission Executables

- Planning
  - Long Term Planning
  - Apportionment/Planning
  - Crypto Planning
  - Payload Table Gen

- Resource Monitor
  - Msip Ops Data Capture
  - CINC Distribution
  - Resource Monitoring
  - TOR

- Guard Subsystem
  - Terminal Data Guard
  - Terminal Control Guard

- Terminal Support
  - Terminal Data Generation
  - Terminal Control PTP

- MPE
  - MPE Mega-Executables
  - MPE Element Spec
  - Component Requirement Specifications

- IP
- AKW/NSA
- MOPS
- OSSE

Many Requirements Spanning Multiple Executables
### MPE Software Architecture

<table>
<thead>
<tr>
<th>Layer</th>
<th>Description</th>
<th>Technology</th>
<th>Access</th>
</tr>
</thead>
<tbody>
<tr>
<td>HMI</td>
<td>User Interface</td>
<td>C/C++ MFC</td>
<td>Presentation / Mission Layer</td>
</tr>
<tr>
<td>Business Rules</td>
<td>Logic</td>
<td>C/C++ SQL</td>
<td>Functional DLL</td>
</tr>
<tr>
<td>Database</td>
<td>Storage</td>
<td>C/C++ SQL</td>
<td>Oracle</td>
</tr>
<tr>
<td>Windows XP</td>
<td>Operating System</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Diagram: Business Rules Span Multiple Layers
How should you test a system like this?
Approach and Challenges

- It’s Not the Same as Last Time
MPE Challenges

- Organizational Bias
  - Bias towards doing the way we did it last time
  - Desire to reuse previous standard test procedures

- Requirements Definition
  - MPE expected to be “all things for all users”
  - Requirements are never perfect
  - Evolving ConOps (and requirements interpretation)

- Requirements Verification
  - Requirements sold-off at lowest applicable level of testing
    - Unintended Regression Danger
  - Significant effort for testing of scenarios and DR removal remain after many requirements already “sold off”

- Schedule Pressure
  - Focus getting cases completed, not straying from the script
  - Testing this system difficult and time consuming
Test Scripts Are Not Correct
- Can’t always capture the steps needed to prove the software was working correctly all the time

Inattentinal Blindness
- Blind to errors that are not part of the test case

Inaccurate Perceptions
- Perception that completing test cases is, by itself, an accurate measure of progress – it is not
- Perception that if you have passed the test cases the software is ready
Observations and Conclusions

- **Overall**
  - Product acceptance based on requirement verification alone not the correct strategy for a product as complex and flexible as MPE

- **Specific**
  - Formal tests typically use small input data sets
    - Testers stick to script, many operational logic paths left untested, lack of operational flavor in test
  - Insufficient system–level and operational environment experience among testers
    - Again, many operational logic paths not tested
  - Insufficient peer reviews
    - Defect escapes from design inspections, code inspections
  - Insufficient integration tests
    - Lack of diversity in testing
## Test Process Improvements

<table>
<thead>
<tr>
<th>Challenge</th>
<th>Corrective Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lack of operational oriented testing</td>
<td>• Added Independent Pre–Ops Test program</td>
</tr>
<tr>
<td>Peer reviews and integration and formal test approach insufficient</td>
<td>• Added Formal Integration Test (FIT)</td>
</tr>
<tr>
<td></td>
<td>• Increased Peer Reviews</td>
</tr>
<tr>
<td>Requirements verification alone insufficient</td>
<td>• SPO acceptance after successfully completing government provided Day–in–the–Life (DITL) test</td>
</tr>
</tbody>
</table>
Unscripted Test

- Unscripted, not Unplanned
Classic Software Lifecycle Model

IEEE 12207
MPE Equivalent Lifecycle Model

- Operational Requirements
- System Requirements
- System Design
- Software System Design
- Architecture Based Design™

- Detailed Design
- Detailed Design

- Next Increment

- CUT

- Code & Unit Test

- PM, CM, SQA

- Code & Unit Test

- SW Integration

- Integration Test

- Unit Test

- Mega-Level Test*

- Segment Level Test*

- Element Level Test*

- Inter-Segment Test

- Development & Test

- IST

- SLT

- ELT

- MLT

- RTO, OUE

- Inter-Segment Test

* Requirements Verification
Formal Test Program

Detailed Design ➔ Code & Unit Test ➔ Mega Level Test ➔ Element Level Test ➔ Segment Level Test

Requirements Verification

Test Cases Completed

Earned Value Claimed
Unscripted Testing

- Form of quality checking that does not rely on test scripts. A tester is let loose on the system and encouraged to report all issues.

- How did we incorporate unscripted test in AEHF?
  - Formal Integration Test (FIT)
    - Contractor led mega-level integration test, freedom to test wide range of values
  - Pre-Ops Test
    - Independent test team focused on mission-based unscripted testing
  - Day-in-the-Life Test
    - Three day test written by government for contractor to assess product operability and effectiveness
Formal Integration Test (FIT)

- Developer-led test
- Bridges gap between requirements testing and unscripted testing. Specifies fields to test, but freedom to test wide range of values.
- Matrices developed by examining fields and subset of variables to be tested
- Sequence of parameterized steps associated with a particular function. Variable of Interest x Range of Interest
Why Formal Integration Testing?

- Product experienced latent defect discovery in deployment and acceptance testing
- Process analysis highlighted need to enhance procedure driven requirement–based testing with scripted free–form combinatorial testing

![Expected Defect Density by Phase](image)

Expected: Low Defect Density in Deployment Phase

![Observed Defect Density by Phase](image)

Actual: High (latent) Defect Density in Deployment Phase
Finding the Right fit for FIT

- Formal Integration Testing (FIT) was used across three increments
- The benefit of FIT increased as it moved left, driving out problems earlier in the development lifecycle
Value of Formal Integration Testing

- Reduction in cost: Industry research shows that problems are easier (and cheaper) to fix earlier in the lifecycle
- Reduction in schedule: Fewer problems = fewer test re-starts = shorter test program duration
- Increase in confidence: Higher quality product delivered with fewer defects and fewer required workarounds
What is Pre-Ops Testing?

- Independent test team (non-developers)
- Mission-based unscripted testing
- Mission/domain experts
- Resembles a Beta Program
Hand-Holding vs. Pointing in Right Direction

- **Test Engineer**

  **Formal Test Case**
  Create a communications plan and add the set of service configurations, BSMRCA Events, and Beam Events to that communications plan.
  - On the BSMRCA Events Tab, add the following Event to the Event List:
    - Sat: 10
    - Beam: BSMRCA-1
    - Duty Cycle: 75, 25
    - Date/Time: Start time is default, start of Communications Events.

  **Motivation:** Complete test cases for verification on or before schedule

- **Functional Experts**

  **Pre-Ops Test Objective**
  Plan for a total force deployment (100 terminals, 40 communication services)
  - Plan for a total force deployment (100 terminals, 40 communication services) in a common Comm Plan using all Service types (combination of LDR/MDR/XDR to include OTADD & Navy RB Nets)

  **Motivation:** Find as many defects as possible (within schedule)
Value of Pre-Ops Testing

- **Mission Ops-Based Focus**
  - Ops-like users performing real-world mission tasks
  - Accounts for multiple missions, users and levels of experience and knowledge

- **Similar to typical “beta” testing**
  - Users using product in unanticipated ways
  - Delivers more diversity in testing
  - Uncovers “user experience” issues

- **Finds defects in development, not in operations**
Put Your Money Where Your Mouth Is
Issue

- Requirements verification alone inadequate for proving operability and effectiveness of the MPE product
  - Lack of diversity in testing
  - Users finding defects that should have been found earlier
  - Users finding issues in real-world conditions not represented in factory test environment
DITL Objective

- The primary objective of the DITL was to assess the operability and effectiveness of the MPE product
  - Can it accomplish a real-world mission under real-world conditions?
  - Is it ready for the user?
DITL Test Case Description

Excerpt from Segment Level Test Plan

1.1.1.1 (U) Test Case I7-SLT-S007-003

(U) Days-in-the-Life MPE|Planning

1.1.1.1.1 (U) Test Case Description

(U) The objective of the Increment 7 MPE DITL is to assess operability and effectiveness of the increment 7 MPE product. It will be run once all other Increment 7 requirements and product integration testing are completed. The test will be operationally relevant and exercise a broad array of MPSS capabilities consistent with the Increment 7 functional baseline. **There will be no traditional test script driving the activities at the defined MPSS stations. Instead, MPSS operators execute a set of Government provided tasks nominally expected to occur in operations.** These tasks may span multiple MPSS User Roles and may or may not need to occur sequentially. Each task will be sized appropriately with respect to nominal operational processing and the total effort of all tasks is not expected to exceed 72 hours of test conduct.

- The Contractor will request the MPE DITL test tasks after completing all predecessor test cases to the MPE DITL and baselining all MPSS DITL impacting DRs.

(U) There are no requirements that will be verified in this test case
DITL Characteristics

- Operational Configuration
- Operational Inputs/Outputs
- Contractor personnel as users
- Mission Driven/Mission Scenario
- Contractor does not know ahead of time what will be on the test
- Passing DITL is an exit criteria for Segment Level Test
- Must be able to accomplish the mission
- Operationally acceptable workarounds OK
## DITL Timeline

<table>
<thead>
<tr>
<th>Timeline</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>– 90 Days</td>
<td>Baseline TEM. Agree on starting DB and content, unique test configuration requests</td>
</tr>
<tr>
<td>– 30 Days</td>
<td>Walk Through Event. Multi-Day Site Walk-through of anticipated procedures, test threads, data capture methods, etc.</td>
</tr>
<tr>
<td>– 2 Weeks</td>
<td>Contractor Indicates “Ready To Execute”</td>
</tr>
<tr>
<td>– 3 Days</td>
<td>Contractor Requests MPE DITL Test Scenario and Government confirms entry criteria met (Successful TRR)</td>
</tr>
<tr>
<td>– 3 Days</td>
<td>Government provides DITL Test scenario and supporting data. The government and a contractor representative will review the mission planning scenario to ensure the missions are understood, in scope and relevant to the test objectives.</td>
</tr>
<tr>
<td>+ 0 Days</td>
<td>Execution of DITL scenario steps begins. (End of day tag-ups: (1) Contractor, (2) User Representatives)</td>
</tr>
<tr>
<td>+ 8 Days</td>
<td>Execution completes after 72 hours of test execution time or successful completion of steps</td>
</tr>
<tr>
<td>+ 10 Days</td>
<td>Complete Analysis and Government Chief Software Engineer issues findings (Pass/Fail)</td>
</tr>
</tbody>
</table>
DITL Operational Configuration

- Strat Planning 1
- Strat Planning 2
- Crypto
- MODC
- PL, US/IP Appt Planning MPE
- IP Flow Control MPE
- MPE Archive to OSSE
- International Partner (IP) Planning UCC
- Tactical Planning UCC

<table>
<thead>
<tr>
<th>USER ROLES</th>
<th>Strat1</th>
<th>Crypto</th>
<th>PL, US/IP Appt Planner</th>
<th>Strat2</th>
<th>Crypto</th>
<th>Tactical Planner &amp; DCA*</th>
<th>4SOPS RM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hardware</td>
<td>MPE1</td>
<td>MPE2</td>
<td>MPE3</td>
<td>MPE4</td>
<td>MPE2</td>
<td>MPE5</td>
<td>MPE6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>MPE1</td>
</tr>
<tr>
<td>Mission Threads</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Day 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Migration</td>
<td></td>
</tr>
<tr>
<td><strong>Day 2</strong></td>
<td></td>
<td></td>
<td>Strategic Planning And Operations</td>
<td>Nominal Planning</td>
<td></td>
<td>Nominal Resource Monitoring</td>
<td></td>
</tr>
<tr>
<td><strong>Day 3</strong></td>
<td></td>
<td></td>
<td>Implementation And Distribution</td>
<td>IP Planning</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## DITL Input Scenario Example

### Table 1 – Strategic Planning Tasks (9.5 hours)

<table>
<thead>
<tr>
<th>Task ID</th>
<th>Scenario</th>
<th>Data</th>
<th>Eval Criteria</th>
<th>Output Product to PDF and save to directory (e.g. DITLProducts/MMYYYY)</th>
<th>Time Est.</th>
</tr>
</thead>
<tbody>
<tr>
<td>25.</td>
<td>IMAGE</td>
<td>In Endurance Mission Re-plan, fail M1</td>
<td>Take screenshots of impacted Services Terminals HMI</td>
<td>Comm Plan Summary Report, Terminal and Service Configuration Reports for “Impacted Services Terminals”</td>
<td>1 hr</td>
</tr>
<tr>
<td>26.</td>
<td>IMAGE</td>
<td>The following tasks will be performed to verify functions independently, but are not dependent on previous tasks for strategic planning</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27.</td>
<td>IMAGE</td>
<td>Move M4 sat in position for A3</td>
<td>Reference Tab S-8</td>
<td>Sat moved properly</td>
<td>N/A</td>
</tr>
<tr>
<td>28.</td>
<td>IMAGE</td>
<td>Add A3 in Late Transition mode</td>
<td>Reference Tab S-9</td>
<td>Sat created &amp; payload(s) configured properly</td>
<td>20 min</td>
</tr>
<tr>
<td>29.</td>
<td>IMAGE</td>
<td>Configure crosslinks</td>
<td>Reference Tab S-9</td>
<td>All sat cross-linked successfully</td>
<td>15 min</td>
</tr>
<tr>
<td>30.</td>
<td>IMAGE</td>
<td>Generate A3 Payload Table (using Payload Table Generation Application)</td>
<td>AEHF DCID 51 (Downlink Channel Parameters Table)</td>
<td>Payload Configuration reports</td>
<td></td>
</tr>
<tr>
<td>31.</td>
<td>IMAGE</td>
<td>Make A3 changes in Frequency Planning Wizard</td>
<td>Reference Tab S-10</td>
<td>LDR/MDR/XDR Constellation Configuration Reports</td>
<td></td>
</tr>
<tr>
<td>32.</td>
<td>IMAGE</td>
<td>Save changes to A3 frequency plan as “Late Transition Modified” template</td>
<td>N/A</td>
<td>XDR Frequency Planning Report</td>
<td>10 min</td>
</tr>
</tbody>
</table>
Value of DITL

- Provided a test gate based on proving the software could accomplish a real-world mission
- Gave the government leverage to assess suitability for ops despite no formal suitability requirement
- Incentivized the contractor to introduce new forms of mission based testing using real-world operators (Pre–Ops Testing) to ensure they could get through this gate
- Dramatically improved the quality of the product
Incremental Test

- Progress, not Regress
Automate, Automate, Automate

- A system that is ridiculously hard to test is equally ridiculously hard to regression test.
- Automated testing expedites verification and protects against regression.
- Some systems are more difficult to automate than others.
Design for Test

- Implementing automated testing can be simplified/facilitated through design
Deliverable Test Scripts

- Protects against regression
- Improves and expedites verification
- Promotes good architecture and design
- Supports sustainment activities

Contract

Extensible Test Framework & Core Test Scripts
From Increments to Early Drops

Now they really are beta releases

Government Testing

Government Acceptance

Automated Regression & Pre-Ops

Beta Program
Risk Management

- Quality vs. Progress
Measuring Progress vs. Quality

- Progress or Quality?

- Completed test cases may not reflect quality
- Earned Value claimed on “defects closed” rather than “defects remaining open” can be a misleading indicator of progress
Software Maturity Risk

Initial Risk Definition

• If the software products have a significant quantity of DRs (including DRs of a high severity) and/or a large quantity of projected latent DRs, then the ability to execute System Test will be adversely impacted, program schedules could be delayed and the ability to field the software could be adversely impacted.

• Risk Areas:
  • Unplanned integration problems / rework

Mitigation Plans

<table>
<thead>
<tr>
<th>Milestone</th>
<th>Risk Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Complete DR Assessment</td>
<td>R 5L.5C</td>
</tr>
<tr>
<td>2 Weak SW Reliability Detected</td>
<td>R 5L.5C</td>
</tr>
<tr>
<td>3 SW Reliability Growth Confirmed</td>
<td>R 4L.5C</td>
</tr>
<tr>
<td>4 Strong SW Reliability Growth Detected</td>
<td>R 3L.4C</td>
</tr>
<tr>
<td>5 Strong SW Reliability Growth Confirmed</td>
<td>Y 3L.3C</td>
</tr>
<tr>
<td>6 Day-In-The-Life (DitL) Completion</td>
<td>G 2L.3C</td>
</tr>
<tr>
<td>7 Monitor Inc 5 Maturity Through IST (6K-2)</td>
<td>G 1L.3C</td>
</tr>
</tbody>
</table>

Burn Down Schedule

Challenge to measure SW Reliability – hrs/defect, defect density, defect priority, defect backlog

DitL is the final milestone for Green
Managing Risk

- An unscripted, government developed test like the DitL can improve software quality
  - Establishes a mentality with the contractor that the software just has to work!
  - Less about the metrics, more about functionality and suitability
- How do we measure software maturity?
  - Common understanding of how defects are captured/measured
    - Normalized metrics, Intuitive front-end for viewing defects
  - Assuming a constant level of test and fix rate, defect backlog should be decreasing
    - Completing test cases is not a measure of quality
  - Automated regression
The Contract and Award Fee Plan

- Where Rubber Meets Road
The contract and award fee plan set the stage for program execution and, ultimately, the quality of the product delivered.

The following performance criteria encourage schedule over quality:

- **Technical Milestone Performance**
  - Deliver a product that meets requirements

- **Cost Performance**
  - Earned Value Management (maintain good SPI & CPI)
The Contract

- Deliverable Extensible Test Framework & Core Test Scripts (automated regression)
- Unscripted Testing (FIT, Pre-Ops)
- Confidence Demos (User Run Event)
- Day in the Life Test (Scenario Based Testing)
- Award Fee Plan (quality driven)
- Iterations With Early Drops For Government Testing (Beta)
- Prototype Development
- Government Test Feedback
- Normalized, quantitative defect metrics capture and presentation
- Software Quality Risk Management Plan
Final Thoughts

- Deploy Quality
“TEST LIKE YOU FLY”
Summary: Focus On Quality and Suitability

- Bake unscripted testing into the contract
- Include government developed tests as quality gates, eg. DitL
- Consider including deliverable automated test scripts as a part of the contract
- Scrutinize proposed test strategy and make sure that it is appropriate for the system being developed (requirements verification is seldom enough with today’s software)
- Risk Management Plan should provide a quantitative way to measure software quality as a positive trend and a measurable mitigation plan to address deficiencies
- Award the contractor for delivering quality products. Avoid “binary” criteria that encourage schedule over quality
  - NOT GOOD ENOUGH: Deliver a product that meets requirements
  - BETTER: Deliver a product that meets requirements AND exhibits a high level of quality and suitability as qualified by:
    1. Risk mitigation activities (are we effectively identifying and burning down risk?)
    2. Quantitative metrics (DDPs, Hrs/Defect, Real-time Dashboard Views) for trends, not just threshold points
    3. User input from government test and confidence demos
- Do whatever you can to get the user community involved early – design reviews, prototypes, confidence demos and early drops for government testing
- Consider iterations and early drops over formal incremental deliveries. Making something operational is expensive and time consuming
  - You might (will) end up with costly overruns if (when) the schedule slips
Thank You

All trademarks, service marks, and trade names are the property of their respective owners.
Acronym List

- AEHF – Advanced Extremely High Frequency
- CM – Configuration Management
- CPI – Cost Performance Index
- DBCM – Database Consistency Manager
- DDP – Defect Density Profile
- DITL – Day in the Life
- DLL – Dynamic Link Library
- FIT – Formal Integration Test
- HMI – Human Machine Interface
- HW – Hardware
- IP – International Partner
- LDR – Low Data Rate
- MDR – Medium Data Rate
- MCS – Mission Control Segment
- MFC – Microsoft Foundation Classes
- MODC – Mission Ops Data Capture
- MOPS – Mission Operations Element
- MPE – Mission Planning Element
- MPSS – Mission Planning Sub-System
- ODBC – Open Database Connectivity
- OSSE – Operations Segment Sustainment Element
- OUE – Operational Utility Event
- PM – Program Management
- RSSC – Regional SATCOM Support Center
- RTO – Responsible Test Organization
- SOC – Satellite Operations Center
- SOM – System Operational Manager
- SPI – Schedule Performance Index
- SQA – Software Quality Assurance
- SQL – Structured Query Language
- SW – Software
- UCC – Unified Combat Control
- XDR – eXtended Data Rate