OFFICE OF THE SECRETARY OF DEFENSE
1700 DEFENSE PENTAGON
WASHINGTON, DC 20301-1700

AUG 28 2014

OPERATIONAL TEST
AND EVALUATION

MEMORANDUM FOR UNDER SECRETARY OF DEFENSE FOR ACQUISITION,
TECHNOLOGY AND LOGISTICS

SUBJECT: Air Force Distributed Common Ground System (AF DCGS)

Attached is my operational assessment of AF DCGS based on results from recent testing of Bulk Release 10B, the Air Force's newest increment. It is based on the results of the Force Development Evaluation (FDE) the 605th Test and Evaluation Squadron (605th TES) conducted with Phase 1 at Langley Air Force Base in January 2014 and Phase 2 at Beale Air Force Base in June 2014. Both level II tests were conducted in accordance with a 605th TES-approved test plan. An independent cybersecurity Red Team did not participate in the evaluation.

Bulk Release 10B degraded performance to unacceptable levels, leading to the Air Force's decision to turn off the new AF DCGS applications in order to continue mission operations. Operators found the system difficult to use, and the software does not meet reliability or availability requirements.

The poor results of the recent testing continue a trend of poor test results and indicate programmatic problems dating from 2010 have yet to be addressed. In March 2010, I assessed the AF DCGS Block 10.2 Geospatial Intelligence system to be not effective and not suitable. The Air Force fielded the parts of Block 10.2 that worked, and continued development on the parts that failed to work. In December 2013, the Air Force approved entrance into the recent Bulk Release 10B FDE despite the system having open Category 1 and 2 deficiencies. During this evaluation, we observed issues similar to those reported in my 2010 assessment. The Air Force has repeatedly tried to proceed with the AF DCGS program without first addressing problems discovered during developmental and operational testing.

Based on my test team's and my observations, there are several issues I wish to bring to your attention which reflect an overall lack of maturity and discipline in the acquisition program.

- The program lacks clear performance requirements for planned enhancements; accurate software maturity trend information because of insufficient tracking and reporting procedures; an approved system-engineering plan; and a clear strategy for testing and evaluating program enhancements. It also lacks approved relevant Department of Defense Architectural Framework products.

- Although AF DCGS is large enough to be an Acquisition Category (ACAT) 1 program, it is not considered to be one because in 2009, the Air Force moved the program to the sustainment phase and subsequently divided the program into four ACAT III programs, none of which is large enough to be ACAT 1. The Bulk Release 10B upgrade is one of the four programs. By managing the effort as a

sustainment initiative, the reduced level of oversight and priority is increasing the opportunity for continued problems, lack of resources and priority, and provides a false impression of reduced risk associated with the program.

- In December 2010, you directed the Air Force provide annual reports for AF DCGS as if it were either a Major Defense Acquisition Program or Major Automated Information System program, based on the magnitude of the investment dollars required to sustain AF DCGS.[1] The Air Force reports provide incomplete insight into the four AF DCGS programs and no information on funding spent versus capability delivered or test and evaluation performance.[2]

- A 2012 Red Team of subject matter experts from industry, former Government experience, and academia reviewed the AF DCGS and concluded that the program is not sustainable and the current acquisition strategy is driving the program to a suboptimal and costly state.[3] Their first recommendation was to reassess the current acquisition and execution authority.

To ensure successful operational testing of BR 10B and subsequent releases, I recommend the Air Force take the following actions:

- Proceed to operational testing only when supported by successful development testing, which demonstrates the ability of AF DCGS to operate at anticipated workload levels;

- Complete a cybersecurity assessment of AF DCGS with a certified Red Team, including operationally representative cyber-attacks;

- Document the Air Force's requirements for each delivery for each of the four programs and apply adequate test and evaluation based on the risk assessment;

- Complete the Test and Evaluation Master Plan (TEMP) and submit for DOT&E approval. This TEMP must include an accurate description of the AF DCGS architecture and interfaces sufficient to justify the test approach.

---

[1] *Defense Acquisition Executive Summary (DAES) Meeting Report*, USD(AT&L) Memorandum, December 30, 2010.

[2] *OSD Program Review of AF DCGS*, Air Force Life Cycle Management Center, January 17, 2014.

[3] *AF DCGS Red Team Briefing*, Presented by AF DCGS Red Team, December 2012.

My point of contact for this effort is Mr. Patrick Sul. He may be reached at (571) 372-3809 or at h.p.sul.civ@mail.mil.

J. M. Gilmore
J. Michael Gilmore
Director

Attachment:
As stated

cc:
USD(I)
AF ISR Agency
SAF/AQ
Dir, T&E, HQ USAF
605 TES

Air Force Distributed Common Ground System (AF DCGS)

Summary

This report assesses the Air Force Distributed Common Ground System (AF DCGS) Bulk Release 10B (BR 10B) based on performance during a Force Development Evaluation (FDE). The 605th Test and Evaluation Squadron (605th TES) conducted Phase 1 of the BR 10B FDE from January 14 – 24, 2014, at Distributed Ground Station (DGS) 1 at Langley Air Force Base. They conducted the Phase 2 FDE from June 11 – 19, 2014, at DGS 2 at Beale Air Force Base, California. Both level II tests were conducted in accordance with a 605th TES-approved test plan. An independent cybersecurity Red Team did not participate in the evaluation.

During Phase 1 of the evaluation, AF DCGS BR 10B new software applications caused such significant slowdowns in workflow that the Air Force made the decision not to use the new applications during the second week of Phase 1 testing. Operators used the legacy manual processes throughout the rest of Phase 1 and 2 of the FDE. The system did not meet any of its reliability requirements because of the critical failures and downtime.

While the Distributed Ground Station (DGS) users can execute their missions with AF DCGS, the key enhancements such as External Tasking Service (ETS) and Workflow are not maturing. These new enhancements are expected to be critical for handling the anticipated increase and complexities of sensors and data sources that will be available to users.

A positive value of BR 10B is that it is bringing common baselines across the AF DCGS enterprise. Prior to BR 10B, the Air Force had sites on 10.1 and 10.2 baselines, and some sites were in need of hardware upgrades. This made system modification very hard and expensive. Getting all of the sites on a common baseline should allow for easier updates and modifications of the entire enterprise.

Background

System Description

The AF DCGS provides software tools for operators to task, process, exploit, and disseminate Intelligence, Surveillance, and Reconnaissance information to the Joint Force Air Component Commander. AF DCGS consists of multiple ground systems at dispersed operational sites. AF DCGS participates in the Department of Defense (DoD) intelligence enterprise via the DCGS Integration Backbone, which uses a metadata catalog and discovery service to enable sharing of information among participants (see Figure 1).

High-Level Operational Concept for AF DCGS OV-1
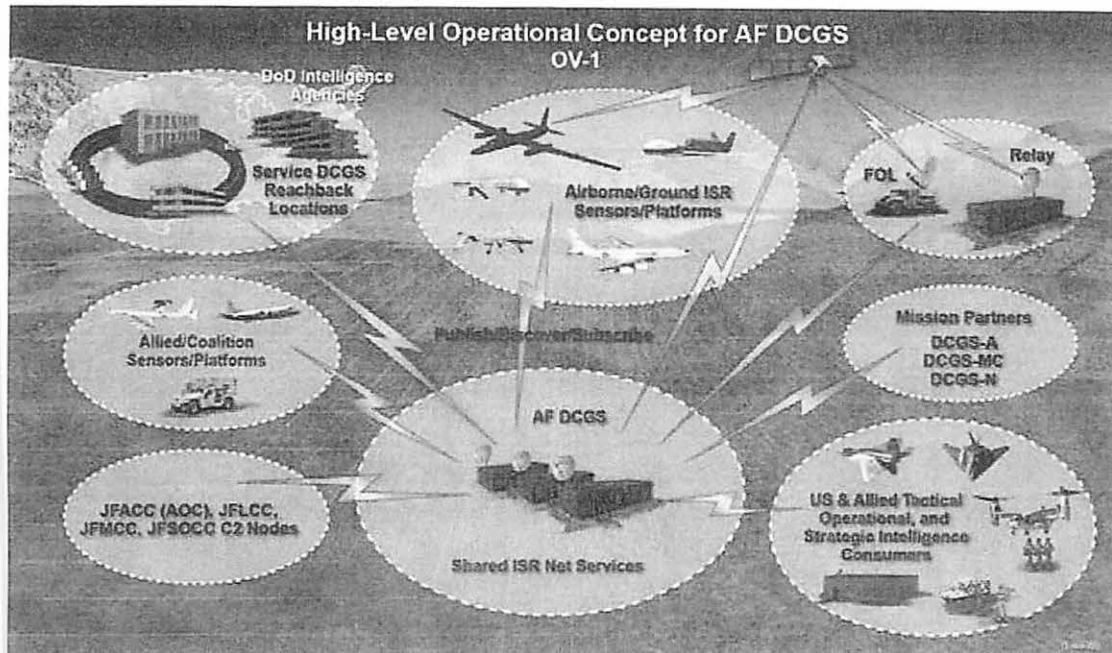
JFACC – Joint Force Air Component Commander
AOC – Air Operations Center
JFLCC- Joint Force Land Component Commander
JFMCC- Joint Force Maritime Component Commander
JFSOCC – Joint Force Special Operations Component Commander
C2 – Command and Control
FOL – Forward Operating Location
ISR – Intelligence, Surveillance, and Reconnaissance
DCGS-A – Distributed Common Ground System – Army
DCGS-M – Distributed Common Ground System – Marine Corps

Figure 1. AF DCGS

The BR 10B replaces portions of the currently fielded Block 10.1 and 10.2 Geospatial Intelligence (GEOINT) baselines to establish a common baseline in all DGSs. BR 10B, as designed, consists of replacement server and router hardware and enhanced software capability in the form of two new applications, Workflow and ETS. Workflow is a software module designed to manage and delegate analyst tasks more efficiently, and ETS is intended to enhance the process of creating and modifying sensor tasks.

AF DCGS Acquisition History and Previous Testing

AFISRA declared Full Operational Capability for AF DCGS effective February 1, 2009. In August 2009, the Assistant Secretary of the Air Force for Acquisition directed the transition of the AF DCGS program to sustainment, despite the fact that much of AF DCGS planned capability had yet to be developed. The AF DCGS program comprises four Acquisition Category III programs: Signal Intelligence (SIGINT) Upgrades, GEOINT Upgrades, Network Communications, and Data Links. The Air Force Operational Test and Evaluation Center declined involvement in all but one of these programs (SIGINT Upgrades), citing that they are only resourced to support test and evaluation for Major Defense Acquisition Programs. The Air Force's goal is to develop AF DCGS through incremental repairs, software upgrades, and end-

of-life hardware equipment replacements (obsolescence). AF DCGS is on DOT&E oversight as a special interest item.

In March 2010, DOT&E evaluated the AF DCGS Block 10.2 GEOINT baseline performance during an FDE, and found that the system was not effective and not suitable. Subsequent to the FDE, AFISRA declared Full Operational Capability and fielded the Block 10.1 with updated hardware from 10.2 along with those 10.2 applications that worked. In December 2013, AFRISA approved entrance into the BR 10B FDE, despite the system having open Category 1 deficiencies and having failed both the developmental and regression test in August and November 2013. As discussed above, during the BR 10B FDE, DOT&E observed issues similar to those reported in the 2010 assessment.

In addition to performance issues observed during tests, programmatic issues have contributed to the inability to institute a successful test strategy. The program lacks adequate system engineering, including an accurate tracking and reporting process for software maturity. The test community did not have access to a common database that captures and prioritizes system problems. This prevents the test community from measuring and categorizing key software metrics such as software problem opening and closure rate and problem duration.

AF DCGS acquisition documents are outdated or have not been created at all. The Test and Evaluation Master Plan (TEMP) is still in draft. The program lacks an updated requirements document that clearly describes the performance requirements for system enhancements, and lacks a system-engineering plan that guides the integration and engineering approach for delivering capabilities. In addition, Department of Defense Architecture Framework documentation is not current.

In December 2010, the Defense Acquisition Executive directed the Air Force to provide annual Program Review reports to him for AF DCGS as if it were either a Major Defense Acquisition Program or Major Automated Information System program, based on the magnitude of the investment dollars required to sustain AF DCGS.[1] The Program Reviews provided by the Air Force provide incomplete insight into the four AF DCGS programs because the reviews do not contain the same type of information required by a Major Automated Information System report.[2] The reviews provide schedules, but no information on funding spent versus capability delivered or test and evaluation performance.

A 2012 Red Team of subject matter experts from industry, former Government experience, and academia reviewed the AF DCGS and concluded that the program is not sustainable and that the current acquisition strategy is driving the program to a suboptimal and costly state.[3] Their first recommendation was to reassess the current acquisition and execution authority.

---

[1] *Defense Acquisition Executive Summary (DAES) Meeting Report*, USD(AT&L) Memorandum, December 30, 2010.

[2] *OSD Program Review of AF DCGS*, Air Force Life Cycle Management Center, January 17, 2014

[3] *AF DCGS Red Team Briefing*, Presented by AF DCGS Red Team, December 2012.

Bulk Release 10B Force Development Evaluation (FDE)

Test Conduct

Bulk Release 10B showed major software shortfalls during both developmental testing in August 2013 and regression testing in November 2013.[4] Despite not meeting the operational test entrance criteria, the Air Force Intelligence, Surveillance, and Reconnaissance Agency (AFISRA) approved entrance into operational testing, waived the entrance criterion of no open Category 1 deficiencies, and deferred the surveillance and warning interface requirement. Based on that guidance, the users turned off Surveillance and Warning DCGS Integration Backbone subscriptions during the operational test.

The 605th Test and Evaluation Squadron (605th TES) conducted Phase 1 of the Bulk Release 10B FDE from January 14 – 24, 2014, in DGS 1 at Langley Air Force Base. They conducted the Phase 2 FDE from June 11 – 19, 2014, in DGS 2 at Beale Air Force Base, California. Both level II tests were conducted in accordance with a 605th TES-approved test plan.[5]

The Phase 1 FDE shadowed 227 hours of real-world missions in DGS 1 and Phase 2 shadowed 144 hours of real-world mission in DGS 2. Data were collected without disrupting current operations and included interviews, questionnaires, recorded observations, and test problem reports. Users were the operators assigned to DGS 1 and DGS 2, and the network and number of internal users were consistent with the network and users intended for fielding.

Test Limitations

Four limitations were known before the test:

- The 605th TES were not able to test the requirement for 1,000 simultaneous internal and external users/connections. The number of internal users was consistent with DGS operations, but the number of external users was lower than expected for high-intensity operations. The test did not determine the ability of users to delegate missions to outside users at other DGSs when the amount of tasking becomes overwhelming.

- The 605th TES did not implement a statistically designed experiment during Phase 1 and Phase 2. Thus, potentially useful quantitative data were not collected and possible factors of interest which may impact performance were not explored. Statistical designs ensure that sufficient quantities of data are collected to support precise and defensible evaluations. Even if all data points required in a design are not collected, as is often the case with real-world observational studies, the portion of data that is lacking can inform the planning for the next phase of testing.

---

[4] *Developmental Test Quick Look Report for Air Force Distributed Common Ground System (AF DCGS) Bulk-Release 10-B (BR-10B) Developmental Test*, 46TS/CZ Memorandum, September 23, 2013.

[5] A Level II test is an "evaluation that includes an independent operational event, which is carried out by typical users in an operationally realistic or representative environment...," *Guidelines for Operational Test and Evaluation of Information and Business Systems*, DOT&E Memorandum, September 14, 2010.

- The 605th TES did not perform a cybersecurity operational vulnerability assessment, as prescribed by DOT&E, with a representative cyber threat Red Team, along with the net defenders participating to evaluate protect, detect, react, and restore functions.

- Finally, the evaluation of joint interoperability was limited to subjective observations because the AF DCGS does not have an Information Support Plan or DoD Architectural Framework products, such as an Operational View-6c, needed for such testing.

The test length was not long enough to prove reliability, availability, and maintainability with sufficiently high statistical confidence. A single-failure test would have required 2,078 testing hours, but only 227 hours were collected during Phase 1 and 144 hours during Phase 2. However, this was not an evaluation limitation for this test because the system had a sufficient number of failures to conclude with high confidence that it did not satisfy the reliability requirements.

Operational Effectiveness

DGS users performed all necessary GEOINT missions, but shortfalls degraded the effectiveness of the system. During the first week of testing in DGS 1, the new web applications, ETS and Workflow, caused unacceptable performance problems and slowdowns. The following is a summary of the Category I problems from Phase 1:[6]

- ETS truncates task priorities, e.g., a priority number of 075 is truncated to 7 while a priority number of 100 is truncated to 1. This is a critical error because truncation can cause relative priorities to reverse, delaying more important tasks.

- ETS removes valid tracks/entities because it erroneously assumes they are duplicates. This forced users to manually correct platform tasking and modify sensor track data.

- Workflow cannot simultaneously claim multiple tracks/entities for the same scene (area of operations), so if multiple images of the same area were requested, only one was actually delivered.

- Workflow slows AF DCGS performance to levels unacceptable to users and causes other functions and applications to freeze. The workaround is to revert to the more cumbersome and slow legacy method of using Microsoft Excel spreadsheets to manually manage tasks.

- AF DCGS could not archive full motion video feeds during the test. Investigation later determined that procedural issues caused the problem and no further issues with full motion video have been reported since the FDE.

The operators turned off ETS and Workflow during the second week of Phase 1 testing because of the errors and inefficiencies caused by the new software. This improved

---

[6] Category I deficiencies are those that can: cause death, severe injury, or severe occupational illness; cause loss or major damage to a weapon system; critically restricts the combat readiness capabilities of the using organization; or result in a production line stoppage. Air Force Technical Manual, *"USAF Deficiency Reporting, Investigation, and Resolution,"* October 1, 2009.

performance, but without ETS and Workflow, Bulk Release 10B provides software capabilities identical to the current 10.1 baseline. Consequently, users still manually managed and assigned tasks, causing inefficiencies in delegating workloads and reducing timeliness of mission completion. With no changes to the Bulk Release 10B software between Phase 1 and Phase 2, operators in Phase 2 also turned off ETS and Workflow off in Phase 2 to avoid the same performance issues.

The legacy 10.1 baseline performance supported operations under normal load conditions from the operators in the DGS, but performance under heavy loads with the required number of simultaneous external users could not be determined. The Red Team study mentioned earlier discussed the growing number of sensors being fielded and the anticipated growth of data entering the DGS that will accompany such advancements. AF DCGS will need the capabilities, which were to have been provided by Bulk Release 10B, to adapt to handle such increased loads, possibly in the near future.

Although the system was observed to be interoperable, the lack of an Information Support Plan and Operational View-6c limited the scope of the evaluation to subjective observations of performance, as testers where provided no information on what interfaces are required.

Operational Suitability

Bulk Release 10B was difficult to use. The 605th Test Squadron administered the System Usability Scale (SUS) survey to all operators, system administrators, maintainers, and original equipment manufacturer field representatives following the Phase 1 FDE.[7] Figure 2 shows the results of the SUS survey. The average score of approximately 45 (80 percent confidence interval of 42.6 to 47.5) is well below the minimum score of 70 for system usability to be considered acceptable.[8] The particularly low scores were likely caused by insufficient training, an inadequate concept of operations, poor tactics, techniques, and procedures, and poor documentation on the system.

Surveys and interviews of the users confirm that Workflow hinders the ability to perform assigned missions and that performance improves without Workflow. Comments also confirm that ETS makes errors and does not interoperate properly with the Unified Collection Operations Reporting Network, known as UNICORN.

---

[7] Phase 2 FDE suitability data and results are not yet available, but observation of the testing and discussion with the test team confirmed that similar results were emerging.

[8] Bangor, Aaron, Philip Kortum, and James Miller. *"Determining what individual SUS scores mean: Adding an adjective rating scale."* Journal of Usability Studies 4.3 (2009): 114-123.
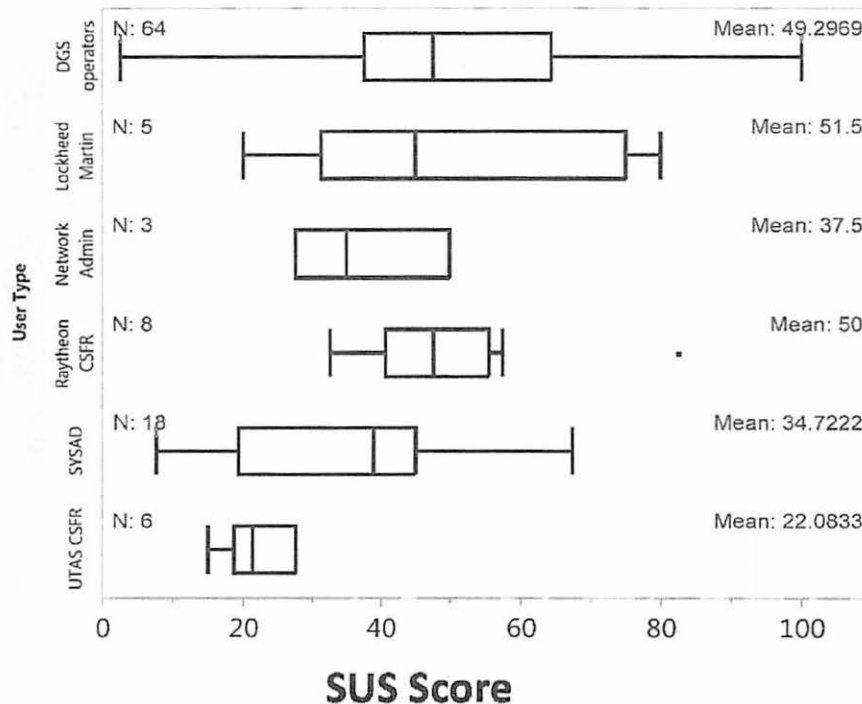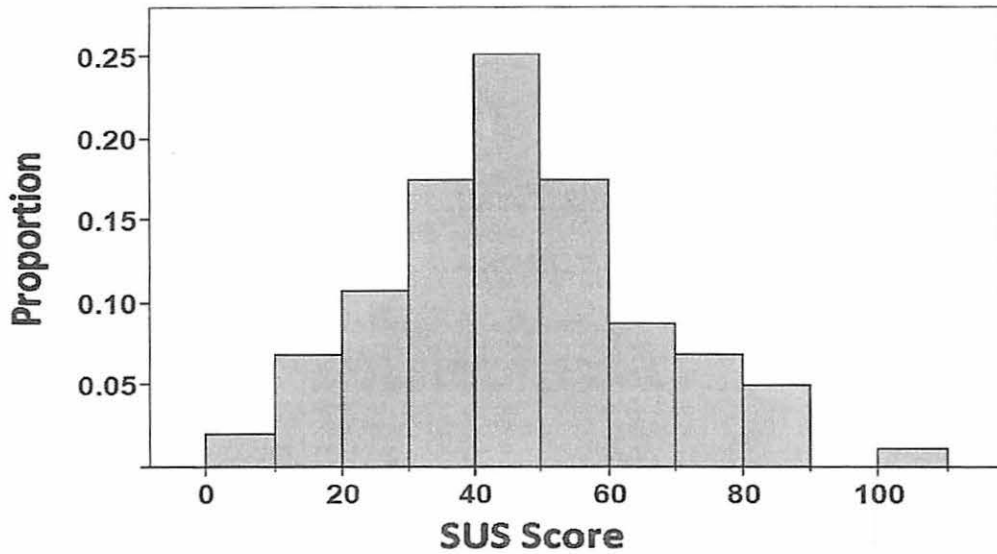
SUS – System Usability; UTAS –United Technologies Aerospace Systems; SYSAD – System Administrator; CSFR – Contractor Support Field Representative; DCS – Distributed Ground Station

Figure 2. System Usability Scale (SUS) scores for all test participants combined (above) and broken down by role (below) for AF DCGS Bulk Release 10B software in the Phase 1 FDE

The BR 10B did not satisfy reliability requirements during Phase 1 FDE. The observed availability of 0.86 (80 percent confidence interval of 0.77 to 0.90) did not satisfy the requirement of 0.9999. The observed Mean Time Between Critical Failure of 16.2 hours (80 percent confidence interval of 12.3 to 21.6 hours) did not meet the requirement of 694 hours. Software applications caused all the critical failures and downtime; no failures or downtime were attributed to the upgraded hardware.

Table 1. Reliability Data and Requirements for AF DCGS Bulk Release 10B during Phase 1 FDE

Measure	Requirement	Phase 1 FDE	80% CI
Availability	95% (All Elements) 99.99% (BR 10B)	0.86 (BR 10B)	(0.77, 0.90)
Mean Time Between Critical Failure	694 hrs	16.2 hrs	(12.3 hrs, 21.6 hrs)
Mean Down Time	<= 1 hr	2.7 hrs	(1.6 hrs, 3.9 hrs)
Mean Repair Time	<= 30 min	2.4 hrs	(1.4 hrs, 3.5 hrs)

FDE – Force Development Evaluation; BR 10B – Bulk Release 10B; CI – Confidence Interval

Reliability results for Phase 2 FDE are not yet available, but are not anticipated to improve. Phase 2 demonstrated 13 reported software problems, including one Category 1 issue regarding intermittent freezing or degradation of full motion video using the Advanced Intelligence Multimedia Exploitation Suite.

Cybersecurity

The AFISRA completed only scans and monitoring activities in support of authority to operate maintenance. AFISRA and the 605th Test Squadron are planning to conduct cybersecurity operational testing in a future event. DOT&E recommends a National Security Agency-certified Red Team conduct the test.

Conclusions

The BR 10B did not successfully deliver the key enhancements expected from this release. New web applications degraded performance to unacceptable levels, causing operators to disable them. The system was difficult to use and did not meet reliability and availability requirements because of the issues with the software.

The results of the recent FDE continue a pattern of unsuccessful developmental and operational tests of AF DCGS. Contributing factors to this pattern include testing before the system is ready, not following through with corrective actions before successive tests, and testing when entrance criteria are not met. The AF DCGS program is not delivering effective capabilities and its activities are consistent with development rather than sustainment. The program lacks the basic foundation of sound acquisition disciplines, including not having in place a rigorous tracking and reporting system for software problems, not having DoD Architectural Framework documents to support interoperability testing, and not having a current and approved TEMP to lay out an overarching test and evaluation strategy.

Recommendations

Based on the Bulk Release 10B FDE and the process leading up to it, DOT&E recommends the following to ensure successful operational testing of BR 10B and subsequent releases:

- Proceed to operational testing only when supported by successful development testing

8

- Demonstrate the ability of AF DCGS to operate at anticipated workload levels
- Complete a cybersecurity assessment with a certified Red Team, including operationally representative cyber-attacks
- Document the Air Force's requiremens for each delivery for each of the four programs and apply adequate test and evaluation based on the risk assessment
- Complete the Test and Evaluation Master Plan and submit for DOT&E approval. This TEMP must include an accurate description of the AF DCGS architecture and interfaces sufficient to justify the test approach

J. Michael Gilmore
Director