The United States Department of Defense (DoD) spends an estimated $20 billion a year on software to support its infrastructure; operate its weapons systems; and provide command, control, communications, computing, intelligence, surveillance, and reconnaissance functions. DoD acquires the large majority of this software from contractor sources.

There have been significant cost overruns and schedule delays experienced in DoD software-intensive system acquisitions, resulting in numerous audits and evaluations of acquisition programs by independent government and industry organizations. Such evaluations have consistently indicated that programs are at risk partially because of failure to implement best practices. The evaluations have recommended the implementation of a variety of practices to improve performance [1-8].

A study by Anderson and Rebentisch [9] of 23 military programs found that practice implementation for eight recommended commercial practices1 ranged from 17 percent to 83 percent with essentially half of the practices implemented under 40 percent of the time and half over 50 percent of the time.

This article reports on research conducted to estimate how broadly acquisition best practices are implemented within DoD. The study involved developing and conducting a survey to establish the implementation and perceived effectiveness of a set of best practices.

**Study Methods**

The first critical issue was deciding whom to survey. It was desirable to obtain as wide a sample as possible with the least amount of interference in the acquisition program activities. For this reason, it was determined that the most effective way to access a wide variety of projects was to contact the various military software centers that provide software expertise to the programs. These centers act as intramilitary consultants or centers of excellence to provide expert resources to the acquisition program offices. Their personnel are in a position to provide informed judgments without any political bias from program loyalty.

```
... evaluations have consistently indicated that programs are at risk partially because of failure to implement best practices.
```

The second critical decision was selecting the best practices for evaluating adoption. Since one of the objectives of this research was to determine how to support the implementation of best practices, it was decided to use the most widely known and oldest set of practices as the baseline for the adoption study. This would maximize the odds that program managers would have heard of the practices and that the acquisition personnel would have encountered them in use. Therefore, the original nine Airlie practices2 were used in the survey.

The following definitions of the Airlie practices are derived from the Software Program Managers Network (SPMN) materials [10, 11]. Copies of the SPMN material defining the practices were included in the materials sent to the survey participants. They are as follows:

1. Formal Risk Management. A formal risk management process requires acceptance of risk as a major consideration; commitment of program resources to managing risk; and use of planned, documented methods for identifying, monitoring, and managing risks.

2. Agreement on Interfaces. A baseline interface specification must be established and agreed to by all stakeholders before implementation activities begin. A separate software specification must be developed with explicit and complete interface information. This is particularly critical with human/machine interfaces and where system interoperability is a requirement.

3. Formal Inspections. Inspections of all acquisition and development documentation should be conducted according to planned, documented processes and the results placed under configuration control, tracked, and resolved.

4. Metrics-Based Scheduling and Management. Statistical quality control of costs and schedules should be maintained. Reasonable cost and schedule projections should be made before program start, and specific measurement processes should be put in place early in the program and rigorously followed. Measurement results should figure prominently in program reviews and management decisions.

5. Binary Quality Gates at the Inch-Pebble Level. Status should be tracked through binary completion of relatively small tasks. Activities are either incomplete or complete. This is to prevent the “80-percent-complete” syndrome where the estimated completion figure is reported without particular rigor.

6. Program-Wide Visibility of Progress vs. Plan. Core indicators of project health and performance should be readily available to all project participants. Anonymous feedback channels...
should be provided to enable bad news to be propagated up and down the project hierarchy without fear of reprisal for truth telling.

7. Defect Tracking Against Quality Targets. Defects should be tracked according to a planned, documented process; measured against established targets; and systematically tracked through removal or resolution.

8. Configuration Management. A planned, documented process is followed to identify, document, monitor, evaluate, control, and approve changes made during the system life cycle to any system-related artifact that is shared by more than one individual or organization.

9. People-Aware Management Accountability. Management should treat personnel as their principle resource by staffing qualified people, encouraging continuous improvement, and fostering an environment conducive to low voluntary personnel turnover.

Survey Instrument
In developing the instrument, we found it useful to think of adoption as having two components: awareness and implementation. Awareness, as defined by Hilburn [12], represents a level of individual knowledge about the practice that includes the following:

- Understanding of the existence and context of the practice within the context of software acquisition.
- A general, informal explanation of the practice.
- Identification of references (human/written) that provides greater depth of knowledge about the practice.

The other component, implementation, requires that the organization put into place the requisite infrastructure, training, resources, and policy to effectively utilize the practice in doing business.

The adoption survey instrument was designed to provide data that addressed both awareness and adoption. It collected data in the following areas:

1. The number of programs supported by the respondent to establish the overall program sample.
2. The size of those programs as designated by Acquisition Category (ACAT): ACAT I, ACAT II, ACAT III, or Other.
3. The quality of practice adoption for each program as measured by the compliance with the practice definition in the Airlie material. This was captured by having each participant differentiate between full compliance and partial compliance for each of the projects. Partial compliance would be a surrogate for awareness, while full compliance would represent implementation. Obviously, this is not a perfect surrogate. However, since partial compliance does, in fact, capture awareness, the worst error would be that of underestimating awareness. It was decided that such an error could be dealt with by considering it in the analysis. The data were gathered for each Airlie practice as the following:

- The number of programs in each size category that fully implemented the practice.
- The number of programs in each size category that implemented some facet of the practice.

4. An evaluation of the perceived overall effectiveness of the practice as observed by the center personnel measured on a five-point scale: Highly Effective (5), Very Effective (4), Moderately Effective (3), Somewhat Effective (2), Negligibly Effective (1). This scale was chosen to reflect that as best practices, the practices were, at worse, ineffective.

Results
Of the 14 centers asked to participate, seven responded to the data call. Six of the seven provided the requested data:

1. Army TACOM TARDEC
2. Army CECOM
3. Navy NAVSEA
4. Navy NUWC
5. Navy NAVAIR
6. Air Force ESC

The seventh respondent, Air Force ASC, provided narrative comments only. The responses covered 150 software acquisition programs broken out as shown in Tables 1 through 3.

Airlie Practice Adoption Data
The responses to the survey resulted in 1,350 possible program-practice pairs (150 programs times nine practices) where a particular Airlie practice could be adopted by a particular program. Table 4 shows the summarized results of the survey when calculated against this full complement of program-practice pairs. The terms partial and full refer to whether the respondent indicated that the program partially or fully implemented the practice.

Table 5 presents the adoption data by practice. The percentages represent the

<table>
<thead>
<tr>
<th>Service</th>
<th>ACAT I</th>
<th>ACAT II</th>
<th>ACAT III</th>
<th>Other</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Air Force</td>
<td>1</td>
<td>0</td>
<td>16</td>
<td>62</td>
<td>79</td>
</tr>
<tr>
<td>Army</td>
<td>9</td>
<td>11</td>
<td>13</td>
<td>6</td>
<td>39</td>
</tr>
<tr>
<td>Navy</td>
<td>2</td>
<td>6</td>
<td>7</td>
<td>17</td>
<td>32</td>
</tr>
<tr>
<td>Total</td>
<td>12</td>
<td>17</td>
<td>36</td>
<td>85</td>
<td>150</td>
</tr>
</tbody>
</table>

Table 1: Programs Reported by Service and ACAT Designation

<table>
<thead>
<tr>
<th>ACAT I</th>
<th>ACAT II</th>
<th>ACAT III</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>8%</td>
<td>11%</td>
<td>24%</td>
<td>57%</td>
</tr>
<tr>
<td>Air Force</td>
<td>53%</td>
<td>26%</td>
<td>21%</td>
</tr>
</tbody>
</table>

Table 3: Percentage of Programs Represented by Service

<table>
<thead>
<tr>
<th>Overall</th>
<th>ACAT I</th>
<th>ACAT II</th>
<th>ACAT III</th>
<th>Other</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Partial</td>
<td>67%</td>
<td>52%</td>
<td>38%</td>
<td>69%</td>
<td>59%</td>
</tr>
<tr>
<td>Full</td>
<td>13%</td>
<td>29%</td>
<td>31%</td>
<td>24%</td>
<td>25%</td>
</tr>
<tr>
<td>Total</td>
<td>80%</td>
<td>80%</td>
<td>69%</td>
<td>93%</td>
<td>84%</td>
</tr>
</tbody>
</table>

Table 4: Overall Results of Adoption Study (Percent of Possible Program/Practice Pairs)
### Table 5: Adoption Results by Practice

<table>
<thead>
<tr>
<th>Best Practice</th>
<th>Partial</th>
<th>Full</th>
<th>Partial</th>
<th>Full</th>
<th>Partial</th>
<th>Full</th>
<th>Partial</th>
<th>Full</th>
<th>Partial</th>
<th>Full</th>
</tr>
</thead>
<tbody>
<tr>
<td>Formal risk management</td>
<td>92%</td>
<td>8%</td>
<td>41%</td>
<td>29%</td>
<td>39%</td>
<td>31%</td>
<td>72%</td>
<td>19%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agreement on interfaces</td>
<td>75%</td>
<td>25%</td>
<td>41%</td>
<td>59%</td>
<td>47%</td>
<td>36%</td>
<td>67%</td>
<td>29%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Formal inspections</td>
<td>50%</td>
<td>8%</td>
<td>71%</td>
<td>0%</td>
<td>36%</td>
<td>28%</td>
<td>65%</td>
<td>20%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metric-based scheduling and management</td>
<td>83%</td>
<td>8%</td>
<td>59%</td>
<td>12%</td>
<td>44%</td>
<td>22%</td>
<td>69%</td>
<td>19%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Binary quality gates at the inch-pebble level</td>
<td>33%</td>
<td>0%</td>
<td>65%</td>
<td>6%</td>
<td>14%</td>
<td>22%</td>
<td>71%</td>
<td>16%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Program-wide visibility of progress vs. plan</td>
<td>83%</td>
<td>8%</td>
<td>47%</td>
<td>53%</td>
<td>47%</td>
<td>36%</td>
<td>72%</td>
<td>26%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Defect tracking against quality targets</td>
<td>25%</td>
<td>17%</td>
<td>29%</td>
<td>41%</td>
<td>31%</td>
<td>36%</td>
<td>68%</td>
<td>31%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Configuration management</td>
<td>67%</td>
<td>33%</td>
<td>47%</td>
<td>53%</td>
<td>42%</td>
<td>42%</td>
<td>66%</td>
<td>34%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>People-aware management accountability</td>
<td>92%</td>
<td>8%</td>
<td>65%</td>
<td>6%</td>
<td>42%</td>
<td>25%</td>
<td>68%</td>
<td>21%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 6: Effectiveness Results by Practice

<table>
<thead>
<tr>
<th>Best Practice</th>
<th>Effectiveness</th>
<th>Effectiveness Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Formal risk management</td>
<td>14%</td>
<td>3.71</td>
</tr>
<tr>
<td>Agreement on interfaces</td>
<td>36%</td>
<td>4.14</td>
</tr>
<tr>
<td>Formal inspections</td>
<td>21%</td>
<td>3.43</td>
</tr>
<tr>
<td>Metric-based scheduling and management</td>
<td>15%</td>
<td>3.69</td>
</tr>
<tr>
<td>Binary quality gates at the inch-pebble level</td>
<td>18%</td>
<td>3.27</td>
</tr>
<tr>
<td>Program-wide visibility of progress vs. plan</td>
<td>15%</td>
<td>3.52</td>
</tr>
<tr>
<td>Defect tracking against quality targets</td>
<td>23%</td>
<td>3.85</td>
</tr>
<tr>
<td>Configuration management</td>
<td>33%</td>
<td>4.13</td>
</tr>
<tr>
<td>People-aware management accountability</td>
<td>14%</td>
<td>3.29</td>
</tr>
</tbody>
</table>

*HE = Highly Effective, VE = Very Effective, ME = Moderately Effective, SE = Somewhat Effective, NE = Negligibly Effective.

### Table 7: Overall Adoption and Effectiveness (Practices in Order of Effectiveness Ranking)

<table>
<thead>
<tr>
<th>Overall</th>
<th>P+F*</th>
<th>F Only</th>
<th>Effectiveness**</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agreement on interfaces</td>
<td>94%</td>
<td>34%</td>
<td>4.14</td>
</tr>
<tr>
<td>Configuration management</td>
<td>96%</td>
<td>38%</td>
<td>4.13</td>
</tr>
<tr>
<td>Program wide visibility of progress vs. plan</td>
<td>94%</td>
<td>30%</td>
<td>3.92</td>
</tr>
<tr>
<td>Defect tracking against quality targets</td>
<td>83%</td>
<td>32%</td>
<td>3.85</td>
</tr>
<tr>
<td>Formal risk management</td>
<td>84%</td>
<td>22%</td>
<td>3.71</td>
</tr>
<tr>
<td>Metric-based scheduling and management</td>
<td>81%</td>
<td>18%</td>
<td>3.69</td>
</tr>
<tr>
<td>Formal inspections</td>
<td>76%</td>
<td>19%</td>
<td>3.43</td>
</tr>
<tr>
<td>People-aware management accountability</td>
<td>83%</td>
<td>19%</td>
<td>3.29</td>
</tr>
<tr>
<td>Binary quality gates at the inch-pebble level</td>
<td>69%</td>
<td>15%</td>
<td>3.27</td>
</tr>
</tbody>
</table>

*P+F* = Partial and full implementations combined; F Only = Full implementations only.
**HE = Highly Effective, VE = Very Effective, ME = Moderately Effective, SE = Somewhat Effective, NE = Negligibly Effective.

Common barriers reported were as follows:
- Lack of Resources: In general, program offices lack sufficient software-educated staff to implement many of the practices. Some lack sufficient core staff and are too busy fighting everyday fires to even consider best practices.
- Inadequate Contracts: Some contracts do not allow the flexibility needed to implement practices that require developer support without costly time- and energy-consuming modifications.
- Data Accuracy, Availability, and Latency: Practices often depend on accurate, timely data that are simply not available to many DoD programs.

### Analysis

The first observation is that the practices are widely recognized across the programs as indicated by the high percentage (84 percent) of either partial or full implementation. Any particular practice was implemented in some form for at least 69 percent of the projects reporting. Three of the practices were implemented in around 95 percent of the projects. However, when only considering full implementations, those figures drop dramatically to an average of 25 percent across all programs with the lowest rate for a specific practice adoption of 15 percent and the highest rate of 38 percent.

Across services, adoption rates were generally consistent. The Army had a generally higher percentage of full implementation. Figure 1 illustrates the relationships between the services. The practices were evaluated as relatively effective, with the majority of responses falling in the moderately effective to very effective range. The practice considered least effective (binary quality gates) still had a mean score of 3.27, placing it in the moderately effective category.

If we look at the adoption and effectiveness data together, there is some correlation between the perceived effectiveness of a practice and its adoption. This seems logical, since in budget- and schedule-constrained programs, the practices with the highest effectiveness would seem more likely to be implemented.

### Conclusions

As stated earlier, we have used partial compliance as a surrogate that implies awareness and full compliance to imply implementation. Therefore, the primary finding from the adoption research is that despite the widespread awareness of the best practices (the average of programs implementing practices was 85 percent, which as we
percent Projects Reporting Practice Use

![Percent Projects Reporting Practice Use](image-url)

- Navy any
- Air Force full
- Navy full
- Army any
- Air Force any
- Army full

Figure 1: Comparison of Service Adoption Rates

of describing practices in a way more suitable to selection and evaluation by acquisition personnel is in the final stages. The results will be briefed in the SIS track during the 2002 Software Technology Conference. The best practice research will be combined with other SIS efforts, including the TriService Assessment Initiative and the CeBASE Experience Factory pilots, to support better decision making and improved processes in software-intensive system acquisitions across DoD. Further work on documenting and disseminating best practices is being performed in collaboration with the Data and Analysis Center for Software in Rome, N.Y.

References


Notes
1. The eight practices studied were past performance, best value, commercial warranties, government/contractor relationship, performance-based specifications, commercial specifications and standards, streamlined contract administration, and use of commercial off-the-shelf/non-developmental item components. The study also found considerable benefits with few drawbacks for using the practices.

2. The nine Airlie Practices were established in 1995 by a group of experts convened by the Navy’s Software Program Manager’s Network at the Airlie Center outside Warrenton, Va. (now historically referred to as “The Airlie Council”).

3. ACAT I is defined as an acquisition program that is not a highly sensitive classified program, and that is designated as a Major Defense Acquisition Program; or estimated to require an eventual total expenditure for research, development, test, and evaluation of more than $355 million in fiscal year (FY) 1996 constant dollars; or for procurement of more than $2.135 billion in FY 1996 constant dollars. ACAT II is defined as those acquisition programs that do not meet the criteria for an ACAT I program, but are estimated to require an eventual total expenditure for research, development, test, and evaluation of more than $135 million in FY 1996 constant dollars; or for procurement of more than $640 million in FY 1996 constant dollars; or if designated as major by the DoD component head. ACAT III is defined as those acquisition programs that do not meet the criteria for an ACAT I or ACAT II. Other is defined as any acquisitions not designated with an ACAT level.

About the Author


1931 Jefferson Davis Highway
Ste. 104
Arlington, VA 22202
Phone: (703) 602-0851 ext. 124
E-mail: rich.turner@osd.mil

Call for Articles

If your experience or research has produced information that could be useful to others, CROSSTALK can get the word out. We are especially looking for articles in several specific, high-interest areas. Upcoming issues of CROSSTALK will have special, yet non-exclusive, focuses on the following tentative themes:

Agile Software Development
October 2002
Submission Deadline: May 20, 2002

Programming Languages
November 2002
Submission Deadline: June 17, 2002

Year of the Engineer and Scientist
December 2002
Submission Deadline: July 15, 2002

Back to the Basics
January 2003
Submission Deadline: August 19, 2002

Be sure to visit us at Booth 509 at the Software Technology Conference in Salt Lake City.

Please follow the Author Guidelines for CROSSTALK, available on the Internet at:
We accept article submissions on all software-related topics at any time, along with Open Forum articles, letters to the editor, and BackTalk.