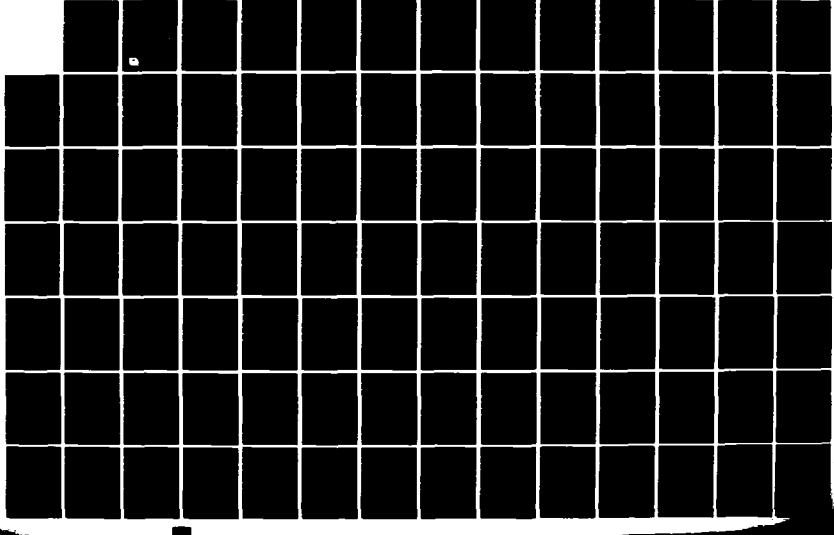
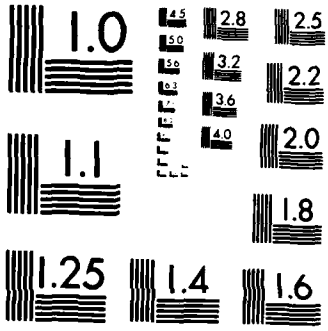
AD-A154 467 ADA (TRADEMARK) PROGRAMMING SUPPORT ENVIRONMENT (APSE) EVALUATION AND VAL.. (U) INSTITUTE FOR DEFENSE ANALYSES ALEXANDRIA VA V L CASTOR ET AL. DEC 84 IDA-M-34 IDA/HQ-84-29295 MDA903-84-C-0031 F/G 9/2 NL

1/2

UNCLASSIFIED





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963-A

AD-E800721
Copy 11 of 83 copies

AD-A156 667

(2)

IDA MEMORANDUM REPORT M-34

Ada* PROGRAMMING SUPPORT ENVIRONMENT (APSE) EVALUATION AND VALIDATION (E&V) WORKSHOP REPORT

Virginia L. Castor
Christine M. Anderson
Timothy E. Lindquist
John F. Kramer, Jr.

DTIC
ELECTE
JUL 12 1985
B

December 1984

DTIC FILE COPY

Prepared for
Office of the Under Secretary of Defense for Research and Engineering

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

INSTITUTE FOR DEFENSE ANALYSES
1801 N. Beauregard Street, Alexandria, Virginia 22311

85 07 09 073

Ada* is a registered trademark of the U.S. Government
(Ada Joint Program Office)

IDA Log No. HQ 84-29295

The work reported in this document was conducted under contract MDA 903 84 C 0031 for the Department of Defense. The publication of this IDA Memorandum Report does not indicate endorsement by the Department of Defense, nor should the contents be construed as reflecting the official position of that agency.

This Memorandum Report is published in order to make available the material it contains for the use and convenience of interested parties. The material has not necessarily been completely evaluated and analyzed, nor subjected to IDA review.

Approved for public release; unlimited distribution.

DISCLAIMER NOTICE

THIS DOCUMENT IS BEST QUALITY PRACTICABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION: Unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY: DD Form 254 dated 1 October 1983	3. DISTRIBUTION/AVAILABILITY OF REPORT: Approved for public release; unlimited distribution
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE: N/A	
4. PERFORMING ORGANIZATION REPORT NUMBER(S): IDA Memorandum Report M- 34	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION: Institute for Defense Analyses; 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION: OUSDRE (DoD IDA Management Office)
6c. ADDRESS (City, State, and ZIP Code): 1801 N. Beauregard Street Alexandria, Virginia 22311	7b. ADDRESS (City, State, and ZIP Code): 1801 N. Beauregard Street Alexandria, Virginia 22311
8a. NAME OF FUNDING/SPONSORING ORGANIZATION: Ada Joint Program Office (OUSDRE); 8b. OFFICE SYMBOL (If applicable): AJPO	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: MDA 903 84 C 0031
8c. ADDRESS (City, State, and ZIP Code): 1211 Fern Street, Room C-107 Arlington, Virginia 22202	10. SOURCE OF FUNDING NUMBERS: PROGRAM ELEMENT NO. / PROJECT NO. / TASK NO. T-4-222 / WORK UNIT ACCESSION NO.

11. TITLE (Include Security Classification): Ada Programming Support Environment (APSE) Evaluation & Validation Workshop Report

12. PERSONAL AUTHOR(S): Christine Anderson, Virginia Castor, John Kramer, Timothy Lindquist

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
FINAL	FROM _____ TO _____	December 1984	123

16. SUPPLEMENTARY NOTATION

17. COSATI CODES		
FIELD	GROUP	SUB-GROUP

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number): evaluation, validation, Ada programming support environment (APSE), requirements, life-cycle support, Ada, Ada Joint Program Office, Common APSE Interface Set (CAIS)

19. ABSTRACT (Continue on reverse if necessary and identify by block number):

The evaluation & Validation (E&V) Task has been established by the AJPO to develop technology necessary to assess the quality of software tools that will be included in an Ada Programming Support Environment (APSE). An E&V Workshop

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT	21. ABSTRACT SECURITY CLASSIFICATION
☐ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	Unclassified

22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE (Include Area Code)	22c. OFFICE SYMBOL

FORM DD 1 JAN 73 1473 84 MAR

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

convened April 2-6, 1984, consisted of government and industry representatives who developed draft docuemnts in three E&V subject areas--Task Recommendations, Task Requirements, and the APSE Evaluation Reference Manual. This report contains an account of the plenary sessions and the papers that address the three subject areas under discussion during this workshop.

IDA MEMORANDUM REPORT M-34

Ada* PROGRAMMING SUPPORT ENVIRONMENT (APSE) EVALUATION AND VALIDATION (E&V) WORKSHOP REPORT

Virginia L. Castor
Christine M. Anderson
Timothy E. Lindquist
John F. Kramer, Jr.

December 1984

DTIC
ELECTE
JUL 12 1985
S D
B

INSTITUTE FOR DEFENSE ANALYSES
1801 N. Beauregard Street
Alexandria, Virginia 22311

Contract MDA 903 84 C 0031
Task T-4-222

Ada* is a registered trademark of the U.S. Government
(Ada Joint Program Office)

ACKNOWLEDGEMENTS

The authors are indebted to all of the Workshop participants for their accomplishments during this labor-intensive week. Because of the expertise and dedication of these individuals, the results of the Workshop have contributed substantially to the overall goals of the DoD E&V effort. Each of these individuals accomplished the difficult task of merging the ideas, suggestions, and concerns of the many industry representatives with varied interests into a cohesive document to support the E&V Task. In addition, we extend our gratitude to the Workshop participants who provided review comments on the rough draft of the Workshop report. These comments have greatly helped improve the quality of the final report. Thanks also go to Bonnie Conover, who assisted in the general planning and arrangements for the Workshop; to Ron Weiner, who assisted in the arrangements at Airlie House; and to Carol Powell, who participated in the final typing of the manuscript. Special appreciation goes to Ms. Anthea DeVaughan, whose enthusiasm, typing skills, and dedication to excellence contributed not only to the effectiveness of the Workshop, but to this report as well.

Accession For
NTIS GRA&I
DTIC TAB
Unannounced
Justification
By
Distribution/
Availability Codes
A-1

PREFACE

The Evaluation and Validation (E&V) Task was established by the Ada Joint Program Office (AJPO) to develop the technology necessary to assess the quality of software tools that will be included in an Ada Programming Support Environment (APSE). The AJPO has designated the Air Force as the lead service for the E&V Task. Virginia L. Castor, AFWAL/AAAF, Wright-Patterson AFB has been designated as chairperson of the E&V Team.

The purpose of the E&V Workshop, convened April 2-6, 1984, at Airlie, Virginia, was to encourage industry participation in the DoD E&V Task and to provide technical input to the E&V team in three areas: recommendations for the E&V Chairperson which address policy, procedures, and technical issues; the definition of requirements for the E&V Task; and development of an approach for preparing an E&V Reference Manual.

The general discussion during the workshop highlighted several areas of concern that could not be addressed fully within the time constraints of this workshop. In particular, the E&V Task must provide near-term relevant and useful technology--"early success"--to the software community and must provide a long-term, all-encompassing view of Ada Programming Support Environments (APSEs) that support integrated life-cycle methodologies which can serve as a framework for organizing E&V documents and activities.

The workshop chairperson was Virginia Castor. She was assisted by three working group chairs: Christine Anderson (AFATL/DLMM, Eglin AFB), Recommendations Working Group; Timothy Lindquist (Virginia Polytechnic Institute and State University), Requirements Working Group; and John (Jack) Kramer (Institute for Defense Analyses), Reference Manual Working Group. This report provides an executive summary of the major workshop accomplishments, an account of the plenary sessions (which inlcuded status reports and final conclusions from each working group), and the completed working group reports.

The accomplishments of this workshop are an important contribution to the E&V task effort. The working group efforts resulted in valuable guidelines for continued teamwork in each area. Moreover, the plenary group discussions

enabled the views from the industry software community to coalesce with those of the DoD community. Many building blocks are needed to complete the E&V task. This workshop provided some of these building blocks.

Table of Contents

Acknowledgements iii

Preface v

Poem ix

1.0 Executive Summary 1-1

1.1 Background 1-1

1.2 Workshop Initiation 1-1

1.3 Workshop Proceedings 1-1

1.4 Organization of This Document 1-3

2.0 E&V Workshop Proceedings 2-1

2.1 Monday, 2 April 1984 2-1

2.1.1 Overview of the E&V Task 2-1

2.1.2 Goals of the E&V Workshop 2-2

2.1.3 Presentation of Position Papers 2-2

2.1.4 Conclusion 2-3

2.2 Tuesday, 3 April 1984 2-4

2.2.1 Goals of the E&V Requirements Working Group 2-4

2.2.2 Goals of the E&V Reference Manual Working Group 2-5

2.2.3 Goals of the E&V Recommendations Working Group 2-6

2.2.4 Conclusion 2-7

2.3 Wednesday, 4 April 1984 2-7

2.3.1 E&V Requirements Working Group Status Report 2-7

2.3.2 E&V Reference Manual Working Group Status Report 2-9

2.3.3 E&V Recommendations Working Group Status Report 2-11

2.3.4 Conclusion 2-14

2.4 Thursday, 5 April 1984 2-14

2.4.1 E&V Recommendations Working Group Final Report 2-14

2.4.2 E&V Requirements Working Group Final Report 2-15

2.4.3 E&V Reference Manual Working Group Final Report 2-15

2.4.4 Conclusion 2-16

2.5 Friday, 6 April 1984 2-16

Appendix A -- Attendees A-1

Appendix B -- E&V Recommendations Working Group Report B-1

Appendix C -- E&V Requirements Working Group Report C-1

Appendix D -- APSE Evaluation Reference Manual Working Group Report D-1

THE APSE E&V WORKSHOP, AIRLIE, VIRGINIA

2-6 APRIL 1984

"Evaluation and Validation"
Has a certain syncopation.

Set to a catchy musical score
This number on the charts might soar.

Just picture those three words in lights
When Jinny sells the movie rights.

We "strange people" came looking for fun
And what did we find? Attila the Hun!

She brought us to a room called "Tack"
And there she put us on the rack.

Assisted by that fiendish pack
Of Chris and Tim and Commander Jack.

They told us of their brutal scheme
To forge a product-oriented team.

They drove us late into the night,
Creating drafts of Papers White.

If high weight is assigned to a life that's hectic,
This environment scored a lofty metric.

They drove our brains to nasty places
Like tool-set coupling and interfaces.

How *should* the "E Team" develop technologies
To evaluate tools and support methodologies?

Should *function* far outweigh performance,
Or ease-of-use, or CAIS conformance?

How should the resulting assessment regime
Relate to a classification scheme?

Why didn't they teach us in Graduate Schools
Of functional taxonomies and canonical tools?

Is validation strictly formal?
Can one think such thoughts and still be normal?

The APSE E&V Workshop
Airlie, Virgina
2-6 April 1984

Throughout the entire four-day history
The pecking order remained a mystery.

The Requirements Folk looked down from high
Upon those lowly Reference guys,

Who promptly solved *their* tribulations
By passing them on to Recommendations.

They, in turn, completed the loop
By dumping on the Requirements Group.

Evaluation is finally clear as a bell:
It's *APSEs* we can't explain very well.

Defining them clearly is *not* elemental.
But the right approach is incremental!

The prize for the snappiest recommendation
Goes to "Federal Bureau of Software Investigation."

Winning the award for hamming and mugging
Are the two guys who did "interactive debugging."

It's a little surprising that it should *turn* out
One copy machine was the only real *burn*-out.

Somehow in the end -- and this may sound funny,
But -- our leader seems more like "Attila the Honey."

She wants to respond to our moans and groans
About flashlights and coffee and room telephones.

So here's to the memory of old Airlie Farm.
We hope that our products don't do any harm.

Our contribution is made -- it can't be deducted.
This workshop is hereby self-destructed.

-- ***The Bard of E&V***
(Bard Crawford, TASC)

1.0 EXECUTIVE SUMMARY

APSE
EVALUATION AND VALIDATION (E&V) WORKSHOP
AIRLIE, VIRGINIA

April 2-6, 1984

1.0 EXECUTIVE SUMMARY

1.1 BACKGROUND

In June 1983 the Ada Joint Program Office (AJPO) proposed the formation of the Ada Programming Support Environment (APSE) Evaluation and Validation (E&V) Task and the establishment of a tri-service APSE E&V Team. The purpose of the E&V Task, in which the Air Force is lead service and the Air Force Wright Aeronautical Laboratories (AFWAL) is lead organization, is to develop the techniques and tools which will provide a capability to perform assessment of APSEs and to determine conformance of APSEs to the AJPO-sponsored Common APSE Interface Set (CAIS) development. As the E&V technology is developed, it will be made available to the community for use by government, industry, and academia.

1.2 WORKSHOP INITIATION

As with previous Ada-related efforts, the participation of representatives from industry and academia is strongly encouraged in the E&V Task. To this end, a Research and Development Sources Sought announcement was published in the Commerce Business Daily (CBD), Issue No. PSA-8503, on January 17, 1984 soliciting the participation of industry representatives in an E&V Workshop in April 1984. Interested firms were requested to respond with the name and resume of their candidate; a brief summary of candidate and corporate history or relevant experience and accomplishments (in areas such as the Ada Program, programming support environments, E&V techniques/methodologies, and life-cycle methodologies); a statement of personal and corporate capability and commitment to participate in the Workshop; and a position paper (4-5 pages) concerning some E&V-related issue. Selection of representatives was based upon the submitted materials.

1.3 WORKSHOP PROCEEDINGS

The E&V Workshop was held April 2-6, 1984, in Airlie, Virginia, for the purpose of identifying E&V APSE-related issues which could then be addressed via the E&V

Task. The goal for this workshop was to develop draft documents in three subject areas:

- E&V Task Recommendations
- E&V Task Requirements
- E&V Reference Manual

The industry representatives, who had been selected on the basis of their responses to the CBD announcement, and members of the E&V Team were organized into subject area working groups. Plenary sessions provided for an exchange of ideas on various subjects within the scope of the workshop and for discussion of each working group's approach to its subject area. Workshop participants expressed a theme of concern that there be "early success" in the E&V task even though there is a need for an all-encompassing E&V technology which supports the implementation of future APSEs. Additionally, the development of a tool/tool-set taxonomy was viewed by all as being essential to provide a framework for all E&V documents and activities.

The Recommendations Working Group, chaired by Christine Anderson, Air Force Armament Laboratory, focused on identifying significant E&V-related issues, categorizing these issues, and formulating recommendations for addressing these issues. The list of issues developed by the group were categorized as policy, procedures, standardization, and other issues that require further consideration before recommendations are formulated.

The Requirements Working Group, chaired by Timothy Lindquist, VPI, organized their activities as steps required to transform the industry input to the E&V Team into technical topics that need to be addressed by a Requirements Document. During the first step, the group reviewed the position papers provided by industry participants and developed three major subject areas that must be addressed by a Requirements Document. These subject areas are:

- Life-Cycle and Methodology Support
- Requirements Based on Application Concerns
- Evaluation of Inter-Tool Interface Complexity

Once the three subject areas were identified, the next step of the Requirements Working Group toward technical topic development required joint sessions with the Reference Manual Working Group. During this joint session, the two groups examined the relationships between the Requirements Document, Reference Manual, and Guidebook, and developed sample entries for these documents. Figure 1 illustrates these relationships. The final step for the Requirements Working Group was to develop an outline for the Requirements Document and to formulate a conceptual framework for presenting APSE evaluation requirements. This conceptual framework includes a phased approach to development of E&V technology and a product quality program.

The Reference Manual Working Group, Chaired by John Kramer, Institute for Defense Analyses, devoted their efforts to establishing an understanding of the relationship among various E&V documents, to defining the users of the Reference Manual and Guidebook, defining the possible uses of the document, and developing first-draft material for various sections of the document.

1.4 ORGANIZATION OF THIS DOCUMENT

Section 2.0 provides the daily proceedings of the Workshop; Appendix A provides the list of attendees; Appendixes B-D contain the reports prepared by each working group.

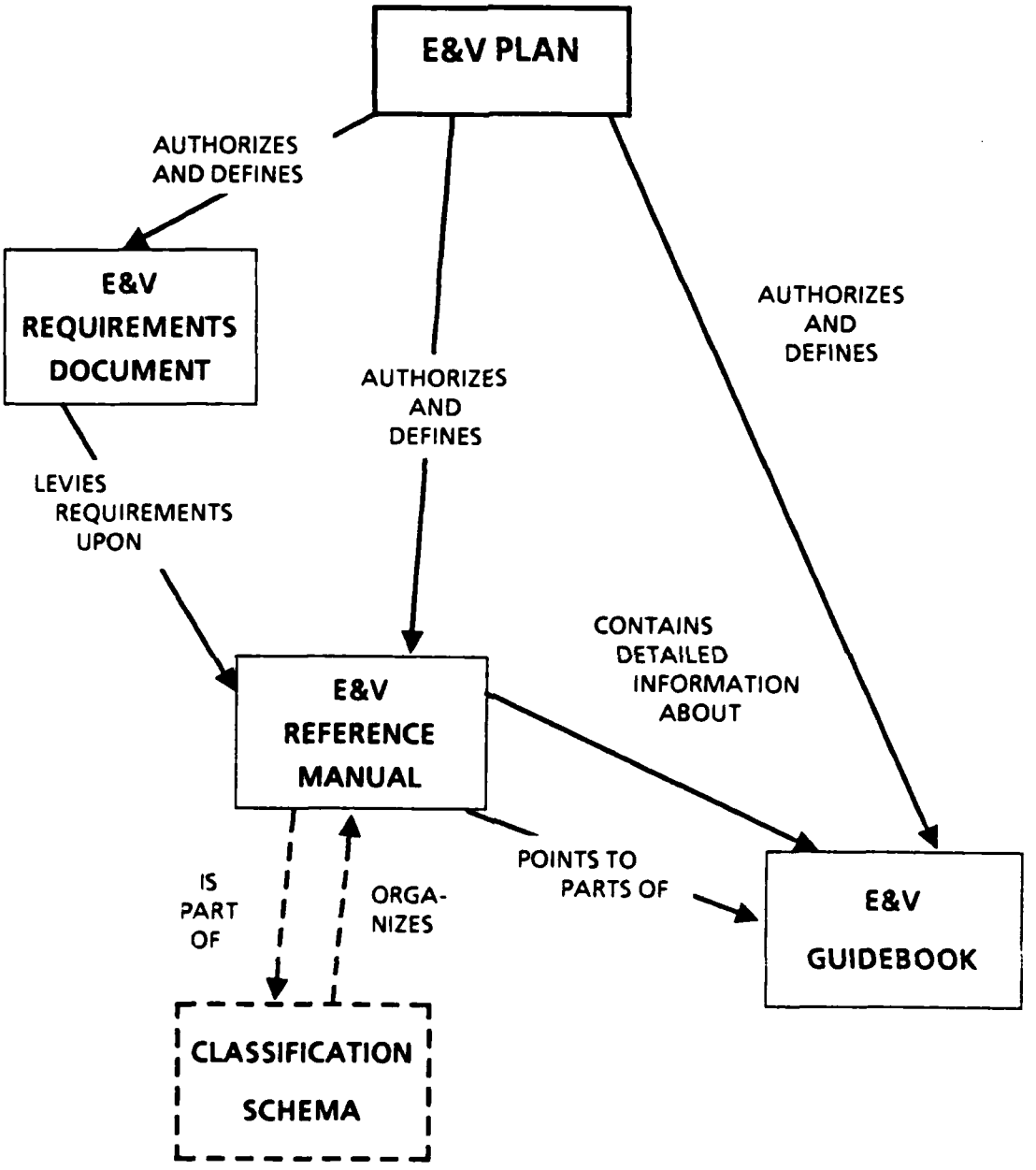


Figure 1. Relationship of Documents

2.0 E&V WORKSHOP PROCEEDINGS

APSE
EVALUATION AND VALIDATION (E&V) WORKSHOP
AIRLIE, VIRGINIA

April 2-6, 1984

2.0 E&V WORKSHOP PROCEEDINGS

This section provides information on discussions and presentations throughout the joint sessions of the workshop.

2.1 MONDAY, 2 APRIL 1984

The E&V Workshop began with registration and a joint meeting of all participants. Following Jinny Castor's welcome to the attendees, each person provided a brief introduction of him/herself. (A complete list of workshop participants is provided in Appendix A.)

2.1.1 Overview of the E&V Task

Each of the industry representatives had been provided a copy of the Evaluation and Validation Plan (Version 1.0 30 November 1983) with the letter of acceptance to the E&V Workshop. Primarily for the benefit of the industry representatives, Jinny provided a brief overview of the E&V Task, which included the following topics:

(a) Concept of an Ada Programming Support Environment (APSE) as defined by the STONEMAN report;
(b) What is meant by Evaluation and Validation (E&V);
(c) Why the E&V Task was established;
(d) Foundation for the E&V Task;
(e) Scope of the E&V Task;
(f) E&V classification schema;
(g) E&V management structure;
(h) E&V Team representation;
(i) E&V Team working groups;
(j) E&V Task relationship to other organizations;
(k) E&V activities to encourage public participation;
(l) E&V schedules for deliverables, meetings and contractual efforts.

2.1.2 Goals of the E&V Workshop

Jinny reiterated that the purpose of the annual E&V Workshop is to encourage industry participation in the E&V Task. She emphasized that participation in future workshops would be limited, with selection of industry participants based upon E&V-related position papers. She outlined that the workshop goals included the development of the following three documents:

(1) Draft E&V Requirements Document (to identify the requirements by which the E&V technology will be developed);

(2) Draft E&V Reference Manual (to provide information on the classification of APSE components);

(3) E&V Recommendations Document (to provide suggestions for procedures, policies, industry/academia involvement, coordination, and related efforts).

2.1.3 Presentations of Position Papers

Each industry representative then provided an overview of his/her position paper which was submitted in response to the CBD announcement soliciting representatives for the workshop. (Actual position papers are not included in this E&V Workshop Report, but will be included as part of the annual E&V Public Report to be published in late 1984.)

(a) Bill Carlson (Intermetrics) -- Some comments on Ada Environment E&V;

(b) Bard Crawford (TASC) -- Simulation: An Important Issue for APSE Evaluation and Validation;

(c) Paul Dobbs (General Dynamics) -- Validation and Standards in Ada Environments;

(d) Bob Fritz (Computer Sciences Corporation) -- Environment Evaluation and Validation: The User's Perspective;

(e) Kathy Gilroy (Harris Corporation) -- Evolutionary Development of an APSE E&V Capability;

(f) Kathy Gracy (SofTech, Inc.) -- Ada Programming Support Environment (APSE) Evaluation Metrics;

(g) Bud Hammons (Texas Instruments) -- APSE Tool Taxonomy;

(h) Asha Kant (Litton Applied Technology) -- Application-specific APSE Evaluation;

(i) Robert Kirkpatrick (Data General) -- Balancing Standardization and Innovation in the Evaluation and Validation of Ada Programming Support Environments;

(j) Mike Meirink (Sperry Corp.) -- APSE E&V Issue: To What Degree Should an APSE Support or Preclude a Specific Genera of Methodology?

(k) Susan Mickel (General Electric) -- Towards APSE Standardization;

(l) Jim Parlier (General Dynamics) -- Standards of Evaluation and Validation for Ada Programming Support Environment Tools;

(m) John Reddan (SYSCON Corp.) -- The Implications of Software Targeted for Embedded Computer Systems on the Evaluation and Validation of Ada Programming Support Environments;

(n) Amos Rohrer (EG&G) -- Compatibility of Run-Time Support for Ada and the CAIS;

(o) Helen Romanowsky (Rockwell International) -- Evaluation and Validation Issues for Embedded Computer Systems Development APSEs;

(p) Andres Rudmik (GTE Communications Systems) -- Increasing APSE Capabilities Impact of E&V;

(q) Raymond Sandborgh (Sperry Corp.) -- APSE E&V Issue: How Can the User Interface be Described in Sufficient Detail and Scope to Perform Systematic Comparisons of APSEs?

(r) Paul Scheffer (Martin Marietta Denver Aerospace) -- Comprehensive Software Development Environments;

(s) Jim Winchester (Hughes Aircraft Co.) -- Evaluating APSE Effectiveness in Developing Software.

2.1.4 Conclusion

During the conclusion of the joint meeting, the issue of DoD E&V-related policies was raised. Jack Kramer pointed out that the development of recommendations for E&V policies and procedures was one of the goals of the E&V Workshop. Jinny concluded by stating that one of the primary reasons why the Air Force accepted lead service responsibility for the E&V effort was that current Air

Force policy regarding the use of APSEs is not as restrictive as that proposed by the Army and Navy. The Air Force expects the E&V Task to provide a near-term capability for the assessment of software tools and environments, regardless of conformance to the Stoneman concept of an APSE. The long-term capabilities must be APSE oriented, which will require further analysis and clarification on the definition of an APSE.

2.2 TUESDAY, 3 APRIL 1984

During this session each of the working group chairpersons described the goals and approaches for that particular working group.

2.2.1 Goals of the E&V Requirements Working Group

Tim Lindquist, Chairperson of the E&V Requirements Working Group, began his presentation by describing the purpose of the E&V Requirements Document, which included:

(a) Define the required approach to the E&V Task;
(b) Define the required E&V Product Quality Guidance;
(c) Define the required APSE componentry to be E&V'd:
    (1) functional taxonomy for APSEs,
    (2) attributes, criteria, standards to be E&V'd against,
    (3) macroscopic APSE considerations.

He stated the activities and anticipated accomplishments of the working group, including:

(a) Discuss the relationship of issues addressed in the industry position papers to the Requirements Document;
(b) Formulate a composite of statements for the E&V Team;
(c) Develop and review sample requirements entries in the Requirements Document and the resulting corresponding entries in the Reference Manual;

(d) Review the initial taxonomy and outline for the Requirements Document as developed by the E&V Team;
(e) Generate inputs to the E&V Team Requirements Working Group.

Tim then presented an interpretation of the relationship between the E&V Requirements Document and the E&V Reference Manual. He stated that the Requirements Document would be used to define the APSE functionality, interfaces and criteria/standards that must be evaluated, and that the Reference Manual would detail how each of the specified requirements would be evaluated or validated. In order to clarify the relationship, Tim provided two examples (the user's interface to a compiler and a compiler's interfaces to other tools) to demonstrate the possible corresponding entries within each document.

2.2.2 Goals of the E&V Reference Manual Working Group

Jack Kramer, Chairperson of the E&V Reference Manual Working Group, began his presentation by describing the relationship of the E&V Reference Manual to other E&V documents. He divided those documents into three categories (predecessor, follow-on, and companion) and provided an overview of each, including the following points:

(1) Closely related predecessor documents:
    (a) E&V Requirements Document,
    (b) E&V Tools/Aids Requirements Document,
    (c) E&V Classification Schema Document;
(2) Closely related follow-on documents:
    (a) APSE Validation Procedures Document,
    (b) CAIS Validation Capability (CVC),
    (c) E&V Tools/Aids;
(3) Companion document:
    (a) E&V Guidebook.

Jack continued by outlining what he hoped would be accomplished by the Working Group:

(a) Define the relationship of the E&V Reference Manual to the E&V Requirements Document (initially addressed by Tim);
(b) Define the contents of the E&V Reference Manual for both the

short term (information useful now) and the long term (when a large volume of information may be detrimental to its effectiveness);

(c) Identify the purpose of the E&V Reference Manual and the audience for whom it is intended.

Jack concluded by listing the deliverables which he expected from the working group sessions:

(a) E&V Reference Manual outline;
(b) Draft contents for each introductory section;
(c) Detail-page format and example;
(d) Solution to the problem of the long-term volume of information and the problem of the length of time which might be required to provide a comprehensive E&V of an APSE;
(e) List of issues identified.

2.2.3 Goals of the E&V Recommendations Working Group

Chris Anderson, Chairperson of the E&V Recommendations Working Group, began her presentation by stating that the objective of the working group was to make recommendations regarding policy and procedure issues associated with the E&V Task. She stated that in order to accomplish that objective the group would initially define the issues, consider alternative approaches, and then formalize the recommendations and include the rationale. She noted that because there was no such equivalent working group on the E&V Team, this Workshop working group provided an excellent opportunity for the industry representatives to submit input to the DoD on policies and procedures on the technical aspects of the E&V Task.

Chris then provided a list of candidate issues extracted from the industry position papers which had been submitted for the Workshop. This list included the following:

(a) Organizational support of E&V;
(b) Public disclosure of results;
(c) Techniques to encourage use of E&V technology;
(d) Yearly validation of APSE components;
(e) Standardization (DIANA, run-time support, APSE descriptive notation, etc.);

(f) E&V technology development schedule (mirror tool availability);
(g) Application-specific benchmarks;
(h) E&V public coordination strategy;
(i) Weighting of E&V tests;
(j) Relationship of E&V to other policies (APSE standardization, etc.).

Chris concluded her presentation by stating that the expected working group deliverable was a Draft E&V Recommendations Document which would consist of the following sections:

(a) Issue identification;
(b) Discussion of alternatives;
(c) Recommendations and rationale.

2.2.4 Conclusion

Following the discussion of the goals of the individual working groups, Jinny Castor reviewed the assignment of individuals to the working groups and the allocation of each working group to a specific room at Airlie.

2.3 WEDNESDAY, 4 APRIL 1984

A joint meeting was held to allow each of the working groups to present a status report of its activities and to enable the other working group members to provide additional inputs to those activities.

2.3.1 E&V Requirements Working Group Status Report

The E&V Requirements Working Group status report was provided by Tim Lindquist, who summarized the activities which had been accomplished by the group.

(a) Brief review of the Requirements Document outline and clarification of the sections;
(b) Definition of common terminology for the group (e.g., "What is an APSE?" and "What is meant by E&V?");

(c) Assumptions used by the group (e.g., APSE definition and interfaces; functional needs such as runtime support and distributed Ada);

(d) Examination of current tool-set criteria (i.e., how well does the current listed criteria take into consideration the interfacing of existing non-APSE and non-CAIS tool-sets such as GKS or EDT?);

(e) Identification of requirements implied by the position papers of the individual working group members:

  (1) An incremental approach to the development of E&V techniques should be used;

  (2) Evaluation of an APSE should distinguish its exclusion/nonexclusion of a methodology;

  (3) Evaluation should include application-specific functionality and efficiency of use;

  (4) Evaluation should include interfaces that support particular embedded computer applications;

  (5) The Requirements Document should include requirements for items which may be deferred in other areas, such as CAIS deferred topics;

  (6) The Requirements Document should include requirements for evaluating the complexity of inter-tool interfaces (tool-set interdependent operation and degree of coupling within a tool-set);

  (7) Two topics were deferred for later inclusion in the Requirements Document outline:

    (i) life-cycle view of E&V,

    (ii) human factors and macroscopic requirements.

Tim concluded by stating that the Requirements Working Group would continue reviewing the white papers prepared by the members, filling in the Requirements Document outline, and working with the Reference Manual Working Group to provide sample corresponding entries to show the relationships between these documents.

2.3.2 E&V Reference Manual Working Group Status Report

The E&V Reference Manual Working Group status report was provided by Jack Kramer, who then provided the following list of deliverables which the group intended to produce:

(a) Reference Manual outline;
(b) Draft contents for each introductory section;
(c) Detail-page format and an example;
(d) A solution to the issues of volume and time associated with the Reference Manual;
(e) A list of identified issues.

Jack continued by identifying the potential users of the Reference Manual. The rationale for this activity is based on the premise that, in order to identify the requirements of the document, one must first identify its potential users:

(a) Tool Procurer (a high-level manager who may not have a technical background);
(b) Tool User (an individual who is competent both technically and as a user of tools on a host machine);
(c) Tool Developer (an individual with software/technical expertise who evaluates the feasibility and viability of a tool);
(d) Quality Assurer (an individual who possesses expertise in the E&V technology and who desires to evaluate APSE tools);
(e) Investor (a potential user of the E&V technology in general as part of a much larger program, rather than with specific tool-sets, i.e., the E&V technology application is embedded within a more global program).

Jack then identified the potential uses of the Reference Manual, based upon the premise that each of the potential users will use the document in a distinct fashion. The document will be used as a:

(a) Technical overview of E&V (an in-depth table of contents providing an introduction and rationale for the E&V effort);
(b) Standards document (to describe the APSE in a legally enforceable and measurable fashion);
(c) Statement of philosophy (a psychological pacifier).

The conclusion reached by the group was that the users (and their characteristics) compose one of the criteria needed for evaluating the content and quality of the manual.

Jack provided an outline for the Reference Manual:

(a) Executive Overview (to present the strategic importance of the E&V effort and documents from the investor perspective);
(b) Scope and Purpose (to present a substantial introduction to the Reference Manual which would include a synopsis of the E&V Plan, a description of the E&V effort, the relationship of the various E&V documents, and a description of the Reference Manual Users;
(c) Applicable Documents (within the context of E&V);
(d) Glossary;
(e) References;
(f) How to Use the Reference Manual;
(g) APSE Global Evaluation (addresses global aspects);
(h) Component Evaluation (addresses tool-specific aspects).

As a further discussion of how the E&V documents are related, Jack presented an example of how compiler E&V issues could be included in each of the documents. He stated that both the Requirements Working Group and the Reference Manual Working Group still had to work together to provide a complete example of the traceability of items through the documents.

Jack then addressed the problems of volume and time associated with using the Reference Manual (i.e., the quantity of information in the manual and the amount of time required to perform a comprehensive evaluation). For the volume problem, the long-term solution will require automated support with possible eventual implementation of an expert system. The time problem can be partially managed by the re-use of environment/tool evaluation results as well as the identification of a user model (view) of the APSE for the purpose of determining the most crucial aspects to be evaluated and validated.

Jack concluded by presenting some of the new issues identified by the working group :

(a) Classifying "hybrid" tools (i.e., tools that are included as parts of other tools)?
(b) Addressing both the global and local aspects of a tool and an APSE.
(c) Renaming the Reference Manual title so that it does not include the term "Validation" (e.g., rename the document to "Evaluation Reference Manual").

2.3.3 E&V Recommendations Working Group Status Report

The E&V Recommendations Working Group status report was provided by Chris Anderson, who began her presentation by summarizing the activities accomplished by the group during the working sessions:

(a) Identification of issues;
(b) Development of Recommendations Document outline;
(c) Discussion of issues;
(d) Assignment of individual write-ups.

Chris stated that the Recommendations Document would comprise four main sections: policy, procedure, standardization, and related issues. Chris then enumerated the topics identified for each of these sections.

Policy-related recommendations included the following:

(a) Organizational support of E&V (OSD-level control organization responsible for application of E&V technology);
(b) Re-E&V of APSE components (required on a periodic basis);
(c) Proprietary source code evaluation (examination of source code for possible impact on maintenance capability);
(d) Public release of test results (maintained by central organization and available as appropriate through NTIS);
(e) Public release of E&V criteria and tests (encourage public awareness and access through NTIS);
(f) Definition of CAIS conformance requirements in the MIL-STD;
(g) DoD directive for CAIS;
(h) Encouraging use of E&V (provide incentive for the application of E&V technology within the Test and Evaluation Master Plan).

The recommendation of evaluation of proprietary information, such as source code and intermediate representations, evoked considerable discussion from the workshop participants. Proponents of the policy identified the need to evaluate whether or not a tool can be reasonably maintained and that such an evaluation would require an examination of the quality of the source code. Opponents of the policy identified legal issues and corporate reluctance to subject proprietary source code to examination. In addition, the opponents argued that examination of the source code would not really enable anyone to judge the degree to which the tool could be maintained.

The presentation of the policy recommendation to define what is meant by CAIS conformance raised the question as to whether such a definition should be specified within the MIL-STD itself, or whether the definition should be provided in a CAIS requirements document, DoD directive, or other more appropriate document.

Chris then presented the procedures-related recommendations which included the following:

(a) Public review of E&V criteria and procedures (provide opportunity for formal review process);

(b) Public coordination with other related efforts and technical groups (encourage inputs from all possible sources);

(c) E&V technology development schedule (mirror tool availability with the development of tool-related E&V capability);

(d) Include weighting of test results in a guidebook used for interpretation of test results (weight test results in order to more easily determine the scope of the test);

(e) Application-specific benchmarks (encourage development and/or contribution of application-specific tests);

(f) Evaluate tool-sets from a functionality perspective as well as a users' perpsective.

The presentation of the procedure recommendation to weight specific tests met with opposition from several workshop participants who indicated that weighting of such tests would be inappropriate for some applications. The concept

of weighting also implies the imposition of decision rules through the application of the E&V technology.

Chris then presented the standardization-related recommendations and noted that the working group had initially addressed several additional items which were later removed from this section and relegated to the other three sections. The standardization-related recommendations included the following:

(a) DIANA (standard intermediate language);

(b) Notation and terminology (standard notation for describing APSE components and standard terminology to be used within the scope of the E&V Task).

The question was raised as to why the Recommendations Working Group had not included text file format standardization on the list. The response was that there were numerous candidates for standardization consideration, but that DoD-imposed standardization for all such candidates was inappropriate and that industry-developed standardization in such areas should be encouraged.

Additional suggestions for consideration by the Recommendations Working Group included inter-tool interfaces as well as an operational definition of an APSE.

Chris concluded with a description of related issues which warrant further consideration:

(a) Maintenance issues (how to measure the ease of maintaining a tool and how to maintain the output of a tool);

(b) Research issues (software metrics, revalidation trigger, automated evaluation tools, standardization of structure, and language-specific benchmarks);

(c) Avoidance of methodology bias;

(d) Legal issues of copyrightable and patentable software;

(e) E&V of mixed-language APSEs;

(f) E&V of tools on different software development architectures;

(g) Run-time support standardization.

2.3.4 Conclusion

During the conclusion of the joint meeting, it was decided that the Requirements Working Group and the Reference Manual Working Group needed to get together during individual working sessions in order to discuss the relationship of the E&V Requirements Document, the E&V Reference Manual, and the E&V Guidebook.

One issue raised during the discussion was that of the scope of the E&V Task and the realization that the volume of tests that would ultimately be generated would be very large. The discussion which followed included comments that the Evaluation aspect of E&V technology application was considerably more labor-intensive than the Validation aspect. It was agreed that the scope issue was significant and that the E&V Reference Manual Working Group was currently attempting to address the volume problem.

The scope of an APSE, as addressed by the E&V Task, was also considered to be a significant issue. Although there was no general consensus as to definition of an APSE, there was agreement that the E&V Task approach to incremental development of E&V technology, based upon existing minimum APSE capabilities, was at least a realistic approach.

2.4 THURSDAY, 5 APRIL 1984

A joint meeting was held in the evening to allow each of the working groups to present a final report on its activities.

2.4.1 E&V Recommendations Working Group Final Report

The E&V Recommendations Working Group final report was provided by Chris Anderson, who stated that the working group had identified issues, provided recommendations, and developed an E&V Recommendations Document (Ref. Appendix D). Chris stated that, except for the following items, there were no changes from her status report given on Wednesday:

(a) Reevaluation was modified to reflect an event-driven schedule rather than periodic intervals, and

(b) The submission of proprietary source code for evaluation purposes would be optional.

2.4.2 E&V Requirements Working Group Final Report

The E&V Requirements Working Group final report was provided by Tim Lindquist, who began his presentation by reviewing the activities which were originally planned for the working group. With regard to the first activity (discuss requirements on E&V which were implied by the position papers), he stated that six white papers in this area had been prepared. With regard to the second activity (develop and review sample requirements and the corresponding entries in the Reference Manual), he stated that the Requirements and Reference Manual Working Groups had met to discuss examples, which would be presented by Jack Kramer. He then reviewed the results of the third activity, an outline of the E&V Requirements Document (Ref. Appendix B).

Tim noted that a white paper was developed which addressed the approach to short-term evaluations with a mapping to long-term (falls under Section 2 of the E&V Requirements Document outline). A white paper was also developed which addressed E&V product quality (falls under Section 3 of the outline). The last major component which was developed addressed the requirements for E&V of an APSE (falls under Section 4 of the outline).

2.4.3 E&V Reference Manual Working Group Final Report

The E&V Reference Manual Working Group final report was provided by Jack Kramer, who stated that the list of issues previously cited had been fully addressed by the group. He reiterated that the group members agreed that the "V" portion of the title of the document should be omitted, yielding "Evaluation Reference Manual". He also stated that the resulting Reference Manual outline had two less sections than previously described in the status report given on Wednesday, a result of better understanding of the problem by the group members.

The first section (Executive Overview) of the Evaluation Reference Manual represents the strategic importance of the effort from the investor's point of view. The second section (Scope and Purpose) reflects the goals of E&V in more detail, the relationship among the E&V documents, and the users/uses of the Evaluation Reference Manual. The third section (Glossary) and the fourth section (References) are self-explanatory, but will require further expansion. The fifth section (How to Use the Reference Manual) provides information regarding efficient usage of the document. The last section (APSE Evaluation) provides detailed information required to conduct an APSE evaluation or component/function evaluation or validation. This section addresses both the near-term considerations in which the Evaluation Reference Manual is represented entirely in document form and the long-term considerations in which much of the detailed information will be automated via use of a data base.

2.4.4 Conclusion

Following the working group final reports, the issue was raised by one of the industry representatives that, although the working groups had accomplished many challenging tasks, the lack of an agreed upon taxonomy to fully represent an APSE caused considerable difficulties for the working groups. After considerable discussion, and due to the late hour, it was agreed that the E&V Team's approach to a taxonomy would be presented to everyone in the wrap-up session the following morning.

2.5 FRIDAY, 6 APRIL 1984

The final joint session of the E&V Workshop was opened by Bard Crawford (TASC), who read a poem which he had written during the week to record the various activities within the Workshop. The poem, which was well received by all of the attendees, is included in the beginning of this document.

Prior to the resumption of the previous evening's taxonomy discussion, Jinny Castor noted that the industry participants in the Workshop would be considered E&V Distinguished Reviewers. Since the formal mechanism by which the Distinguished Reviewers would participate with the E&V Team remained to be

determined, Jinny requested that the workshop participants provide suggestions/recommendations as to how such a mechanism could be established so that she could then coordinate a mechanism through the AJPO.

The discussion of the APSE taxonomy was opened by Tim Lindquist, who addressed the issue of a schedule for the development of the taxonomy. The E&V Team members (Tim Lindquist, Ronnie Martin, and Betsy Bailey) had previously developed an initial approach to the taxonomy based upon phases of the software life cycle. The first version of the taxonomy would be available during the 6-8 June 1984 E&V Team meeting as part of the draft Version 1 E&V Requirements Document. The concern expressed by members of the Reference manual working group was that Version 1 of the E&V Reference Manual, which is due during the first quarter of calendar year 1985, would require a substantially detailed taxonomy on which to base its contents.

A suggestion was made that the initial taxonomy could be incorporated as part of the Workshop report but, because the taxonomy was not a product of the Workshop, the suggestion did not receive approval. Tim polled the industry representatives to determine those who would like to participate in the review of the June 1984 draft Version 1 E&V Requirements Document; he received a favorable response. Tim suggested a review of the draft document by the Distinguished Reviewers during June and July 1984, the results of which could be incorporated in Version 2 of that document. Those comments received during the present discussion would be reviewed and considered for incorporation in Version 1.

The significance of the taxonomy was then addressed. The taxonomy was recognized as the framework upon which all of the E&V documents were based and, therefore, the majority of the E&V Task. Tim then described the initial E&V Team matrix approach to the development of the taxonomy. This approach was based upon software systems life-cycle phases and functions performed within each of those phases. Suggestions from the industry representatives included the use of additional life-cycle phases and the use of a multiply attributed tree representation as additions to the current matrix approach.

Jinny concluded the E&V Workshop by thanking the participants for their efforts and their contributions to the E&V Task. She also requested their continued involvement in the E&V Task as E&V Distinguished Reviewers. The workshop was then formally adjourned.

APPENDIX A

ATTENDEES

APSE
EVALUATION AND VALIDATION (E&V) WORKSHOP
AIRLIE, VIRGINIA

April 2-6, 1984

EVALUATION AND (E&V) WORKSHOP
AIRLIE, VIRGINIA

April 2-6, 1984

ATTENDEES

NAME	TELEPHONE	MAILING ADDRESS	NET ADDRESS
E&V Workshop Chairperson:			
CASTOR, Jinny	(513) 255-2446	AFWAL/AAAF Wright-Patterson AFB OH 45433	CASTOR @ USC-ECLB
E&V Workshop Members:			
ANDERSON, Chris	(904) 882-2961	Air Force Armament Laboratory DLMM Eglin AFB, FL 32542	CANDERSON @ USC-ECLB
BAUER, Lee	(301) 657-3775	Intermetrics, Inc. 4733 Bethesda Avenue Bethesda, MD 20814	
BURTON, Dan	(617) 861-2001	ESD/ALL Hanscom AFB, MA 01731	DBURTON @ ECLB
CARLSON, Bill	(301) 657-3775	Intermetrics, Inc. 4733 Bethesda Avenue Bethesda, MD 20814	BCARLSON @ ECLB
CRAWFORD, Bard	(617) 944-6850	TASC One Jacob Way Reading, MA 01867	
DeVAUGHAN, Anthea	(703) 845-2098	Institute for Defense Analyses 1801 N. Beauregard Street Alexandria, VA 22311	ANTHEA @ USC-ECLB
DOBBS, Paul	(817) 731-0741	General Dynamics Box 748, MZ 5404 Ft Worth, TX 76114	
ESTES, Nelson	(513) 255-5945	Aeronautical Systems Div. Wright-Patterson AFB OH 45424	ESTESN @ WPAFB-JALCF
FRITZ, Robert	(619) 225-8401	Computer Sciences Corp. 4045 Hancock Street San Diego, CA 92110	

E&V Workshop, 2-6 April 1984 -- Attendees

GILROY, Kathy	(395) 676-6363	Harris Corp. GISD, 150 Wickham Road Melbourne, FL 32901	HAIPM @ JPL-VLSI
GRACY, Kathy	(513) 429-2771	SofTech, Inc. 3100 Presidential Drive Fairborn, OH 45324	
HAMMONS, Bud	(214) 462-4690	Texas Instruments, Inc. P.O. Box 405 Lewisville, TX	TIKIT @ USC-ECLB
KANT, Asha	(408) 773-0777 x2164	Litton Applied Technology 645 Almanor Avenue, Bldg. 2 Sunnyvale, CA 94086	
KIRKPATRICK, Robert	(919) 549-8421	Data General Corporation 62 T.W. Alexander Drive Research Triangle Park, NC	
KRAMER, Jack	(703) 845-2263	Institute for Defense Analyses 1801 N. Beauregard Street Alexandria, VA 22311	KRAMER @ USC-ECLB
LINDQUIST, Tim	(703) 961-7537 messages: 6931	Computer Science 562 McBryde Hall Virginia Tech. Blacksburg, VA 24061	LINDQUIST % VPI @ RAND-RELAY
LONG, Rick	(513) 255-2446	AFWAL/AAAF Wright-Patterson AFB OH 45433	RLONG @ USC-ECLB
MARTIN, Ronnie	(404) 894-3837	School of Information and Computer Science Georgia Institute of Technology Atlanta, GA 30332	RONNIE.GATECH @ RAND-RELAY
MEIRINK, Mike	(612) 456-4925	Sperry Corp., CS/DSD P.O. Box 43525 Saint Paul, MN 55165	
MICKEL, Susan	(408) 734-4980	General Electric Co. 1277 Orleans Drive Sunnyvale, CA 94086	
PARLIER, Jim	(619) 576-5241	General Dynamics Data Systems Div. Western Center P.O. Box 85808, MZ: VP 5300 San Diego, CA 92138	

E&V Workshop, 2-6 April 1984 -- Attendees

PROBERT, Thomas H.	(703) 845-2517	Institute for Defense Analyses 1801 N. Beauregard Street Alexandria, VA 22311	PROBERT @ USC-ECLB
REDDAN, John	(619) 296-0085	SYSCON Corp. 3990 Sherman Street San Diego, CA 92110	SCI-ADA @ USC-ISI
ROHRER, Amos	(703) 368-0100	EG&G, WASCI 8809 Sudley Road Manassas, VA 22110	
ROMANOWSKY, Helen	(319) 395-3868	Rockwell International 400 Collins Road N.E. Cedar Rapids, IA 52498	
RUDMIK, Andres	(602) 582-7518	GTE Communications Systems 2500 W. Utopia Road Phoenix, AZ 85027	ARUDMIK @ USC-ECLB
SANDBORGH, Raymond	(612) 456-3185	Sperry Corp., CS/DSD P.O. Box 43525 Saint Paul, MN 55165	
SCHAEFER, Carl	(301) 657-3775	Intermetrics, Inc. 4733 Bethesda Avenue Bethesda, MD 20814	
SCHEFFER, Paul	(303) 977-3605	Martin Marietta Denver Aerospace M/S 0421 P.O. Box 179 Denver, CO 80201	
WILLIAMSON, Jimmy	(513) 255-2446	AFWAL/AAAF-2 Wright-Patterson AFB OH 45433	JWILLIAMSON @ USC-ECLB
WINCHESTER, Jim	(714) 732-5576	Hughes Aircraft Corp. P.O. Box 3310/618-M215 Fullerton, CA 92634	JRUBY @ USC-ECLB

APPENDIX B

E & V RECOMMENDATIONS WORKING GROUP REPORT

APSE
EVALUATION AND VALIDATION (E&V) WORKSHOP
AIRLIE, VIRGINIA

April 2-6, 1984

APPENDIX B

E&V RECOMMENDATIONS WORKING GROUP REPORT

TABLE OF CONTENTS

I.	**E&V RECOMMENDATIONS WORKING GROUP OVERVIEW**		B-3
	1.0	List of Members	B-3
	2.0	Goal	B-4
	3.0	Accomplishments	B-4
II.	**E&V RECOMMENDATIONS DOCUMENT**		B-5
	EXECUTIVE SUMMARY		B-5
	1.0	**INTRODUCTION**	B-8
		1.1 Purpose	B-8
		1.2 Scope	B-8
	2.0	**POLICY ISSUES**	B-9
		2.1 Organizational Support of E&V	B-9
		2.2 Reevaluation and Revalidation (re-E&V) of APSE Components	B-9
		2.3 Proprietary Source Code Evaluation	B-10
		2.4 Public Release of Test Results	B-11
		2.5 Public Release of E&V Criteria and Tests	B-12
		2.6 Definition of "CAIS Conformance"	B-13
		2.7 DoD Directive for CAIS	B-14
		2.8 Encouraging Use of E&V	B-14
	3.0	**PROCEDURE ISSUES**	B-16
		3.1 Public Review of E&V Criteria and Procedures	B-16
		3.2 Public Coordination with Other Related Efforts and Technical Groups	B-17
		3.3 E&V Technology Development Schedule	B-18
		3.4 Weighting of Test Results	B-18
		3.5 Reader's Guide for Interpretation of Test Results	B-19
		3.6 Application-specific Benchmarks	B-20
		3.7 APSE Definition	B-20
	4.0	**STANDARDIZATION ISSUES**	B-22
		4.1 DIANA	B-22
		4.2 APSE Terminology and Notation	B-23

5.0 RELATED ISSUES B-24
5.1 APSE Maintenance Tools B-24
5.2 Research Issues B-24
5.2.1 Metrics Definition B-25
5.2.2 Revalidation Triggers B-25
5.2.3 Inter-Tool Communication B-25
5.2.4 Language Feature-specific Benchmarks B-26
5.2.5 Independent Evaluation of E&V Technology B-26
5.3 Methodology Biases in the Evaluation Process? B-26
5.4 Legal Considerations Relating to E&V B-27
5.5 E&V of Mixed-Language APSEs B-28
5.6 E&V of Distributed APSE Functions B-28
5.7 Run-time Support Standardization B-29

6.0 REFERENCES B-31

I. E&V RECOMMENDATIONS WORKING GROUP OVERVIEW

1.0 RECOMMENDATIONS WORKING GROUP MEMBERS

Chairperson--Christine Anderson, Air Force Armament Laboratory

Members:

Lee Bauer, Intermetrics, Inc.

Robert Fritz, Computer Sciences Corporation

Kathleen Gracy, SofTech, Inc.

Robert Kirkpatrick, Data General Corporation

Susan Mickel, General Electric Company

Amos Rohrer, EG&G Washington Analytical Services Center, Inc.

James Williamson, Air Force Wright Aeronautical Laboratories

2.0 GOAL

The goal was to make recommendations regarding policy and procedure issues associated with the E&V Task.

3.0 ACCOMPLISHMENTS

The working group defined issues, considered alternative approaches, and then formalized the recommendations and included the rationale. Issues and recommendations were categorized as policy, procedures, standardization, and others which require further consideration. Note: The Recommendations Document was reviewed by all E&V Workshop participants, representing 19 industrial organizations.

II. E&V RECOMMENDATIONS DOCUMENT

EXECUTIVE SUMMARY

In June 1983 the Ada Joint Program Office (AJPO) proposed the formation of the Evaluation and Validation (E&V) Task and a tri-service Ada Programming Support Environment (APSE) E&V Team, with the Air Force designated as lead service. In October 1983 the Air Force officially accepted responsibility as lead service on the E&V Task. The purpose of the E&V Task is to develop the techniques and tools which will provide a capability to perform assessment of APSEs and to determine conformance of APSEs to the Common APSE Interface Set (CAIS). The CAIS is a set of interfaces that provides interoperability of data and transportability of tools between conforming APSEs.

In order to encourage industry/academia participation in the E&V effort, an E&V Workshop was held at Airlie, Virginia, from 2-6 April 1984. The following recommendations were made by industry participants at the E&V Workshop:

- The E&V process should be managed by a central facility, possibly the Software Engineering Institute, at the Office of the Secretary of Defense (OSD) level with satellite facilities, possibly subcontractors, in the field to perform the actual testing.

- "Event driven" rather than periodic reevaluation should be performed. Revalidation should be performed on a periodic basis (not less than every two years) or "event driven", whichever comes first. Furthermore, trigger "events" and policies regarding baselining tool versions used on government contractual efforts (including life-cycle support and upgrades) should be explicitly delineated.

- Submission of proprietary source code may be required for certain evaluation tests. This submission should be optional. If the code is not submitted, test results should reflect the fact that the code was not available for review.

- Results of the E&V testing of tools offered for commercial advantage should be maintained in a repository at the central E&V oversight facility and be made available to the public.

- All E&V criteria and tests should be made available to the public.

- "Conformance to the CAIS" should be defined so that a validation suite can be developed by the E&V Team.

- DoD should issue a directive requiring the use of the CAIS when it becomes a military standard and when validation procedures and CAIS implementations are available.

- When the E&V technology matures, evaluation and validation (where appropriate) should be mandated on all government contracts.

- Prior to complete maturation of the technology, certain incentives (short of a mandate) should be provided to encourage public use of E&V tools and techniques. These incentives could include: Director Defense Test and Evaluation (DDT&E) credit given in the Test and Evaluation Management Plan (TEMP) review for E&V'd tools; and consumer reports concerning tool E&V results issued by the central E&V oversight organization.

- The E&V Team should stage a public review of the E&V criteria and procedures prior to final release.

- The E&V Team should coordinate with potential E&V users including non-Ada groups.

- Test results should be weighted so that the scope of an individual test may be more easily understood.

- E&V test results should be accompanied by a test results reader's guide aimed at helping non-Ada, non-software-oriented program managers understand and better use the information.

- Application-specific benchmarks should be collected and developed as part of the E&V Task.

- Refinement of the STONEMAN APSE definition should be performed.

- The following should be standardized: DIANA, APSE component descriptive notation, and E&V terminology.

Other related issues requiring further examination include: E&V liability associated with E&V-ing proprietary software; transition issues involving E&V-ing mixed-language APSEs; E&V-ing APSEs on different types of architectures; and development of automated evaluation tools.

1.0 INTRODUCTION

1.1 Purpose

The purpose of the E&V Recommendations Document is to provide recommendations and identify related issues regarding the E&V Task. These recommendations and issues were formulated by representatives from industry during the E&V Workshop in Airlie, Virginia, from 2-6 April 1984. These recommendations are the views of the participants and are not necessarily organizational positions.

1.2 Scope

The overall objective of the E&V Recommendations Document is to provide a set of detailed recommendations and identify issues that require further examination within the context of the E&V Task. These recommendations and issues apply to the E&V Task, other closely related activities, and deployment of the E&V technology to the public. Accordingly, some of these recommendations can be acted upon by the E&V Team, if deemed appropriate, while others are matters to be confronted by other groups and organizations.

Each recommendation is preceded by a brief description of the issue, the various alternatives for resolving the issue, and the rationale behind the ultimate recommendation.

Related issues are presented more succinctly through identification and brief discussion.

2.0 POLICY ISSUES

This section presents recommendations related to policy issues associated with the E&V Task. Policy issues deal with rules and guidelines for the development, transitioning, and application of the E&V technology.

2.1 Issue: Organizational support of E&V

What type of facility should be established to perform the E&V process?

Discussion: For E&V to have the desired effect, some type of facility must be developed to perform the E&V process. It is recommended that a central E&V facility be established, at the Office of the Secretary of Defense (OSD) level. An advantage of having one high-level responsible office is that all test maintenance and configuration can be performed at one location. This organization could possibly be incorporated into the Software Engineering Institute.

This single facility should have sole responsibility for reviewing, releasing, and maintaining a repository of test results. This facility must be secure in order to perform classified as well as proprietary testing. Additionally, great care must be taken to ensure that proprietary tools are not compromised.

The central E&V facility should have satellite facilities to perform the actual testing of tools. This testing could possibly be accomplished by commercial subcontractors.

Recommendation: **The E&V process should be managed by a central facility, possibly the Software Engineering Institute, at the OSD level with satellite facilities, possibly subcontractors, in the field to perform the actual testing.**

2.2 Issue: Reevaluation and revalidation (re-E&V) of APSE components

What policy should be established regarding reevaluation and revalidation (re-E&V) of APSE components?

Discussion: Considering the expense of validation procedures and the (expected) greater expense of evaluation procedures, it is unreasonable to require re-E&V following each modification to an APSE component. One alternative is to

limit it to "major" modifications. However, the term "major" is open to a variety of interpretations. To avoid possible abuses, a periodic testing requirement should be added. It is especially important for validated tools that revalidation not be postponed for lengthy periods. A combination of the two strategies is recommended for revalidation procedures. APSE tools should be revalidated on a periodic basis, or more frequently based on the occurrence of certain triggering events.

The ACVC plan requires yearly re-validation of Ada compilers. Requiring yearly revalidation of the remaining APSE components is considered prohibitively expensive. Especially with the safeguard of event-triggered revalidation, a significantly longer period of time (at least two years) is recommended.

Reevaluation, however, does not involve conformance to a standard. Therefore, it is not considered necessary to establish a periodic testing requirement. Reevaluation should be entirely event-driven. Most frequently, it is expected that reevaluation will be voluntary based on a desire to improve a tool's quality assessment.

Development of the list of events triggering revalidation and those triggering reevaluation is a research issue which is discussed in a later paragraph. An example of an event that would trigger revalidation is re-hosting on different hardware.

Recommendation: **Revalidation should be required according to a fixed schedule (e.g., every two years) or as required by triggering events (e.g., re-hosting). To contain expense, care should be taken in establishing a reasonable period of time for the fixed schedule. A period considerably longer than one year is suggested. No fixed schedule of reevaluation is considered necessary. Re-evaluation should be entirely event-driven.**

2.3 Issue: Proprietary source code evaluation

Should the submission of proprietary source code be required for evaluation testing?

Discussion: In the E&V process, source code of APSE tools will certainly be examined in the application of various metrics. For example, a complexity metric

may be applied to assess maintainability of the source code. Since source code will be examined, some policy regarding submittal of proprietary source code must be formulated.

Source code for government-procured APSE tools will certainly be evaluated. APSE tools which are intended to be marketed and are evaluated may contain proprietary source code. Tools which are developed for in-house use will contain proprietary code.

In order to promote E&V, the submittal of source code should be encouraged, but not required as a prerequisite. If no source code is submitted with an APSE tool for E&V, then the test results should note that fact by recording it with the other test results. If proprietary source code is submitted for E&V, then appropriate data access and protection agreements must be made by the E&V agency. Data access and protection requirements may necessitate the E&V facility to have a secure system even to the extent to be able to test classified tools.

Recommendation: **Submission of proprietary source code for evaluation tests should be optional. If the code is not submitted, test results should reflect the fact that the code was not available for review. Furthermore, all source code that is submitted must be protected to the extent required.**

2.4 Issue: Public release of test results

How should E&V test results be handled?

Discussion: If E&V test results are to be released to the public, we must also decide if there are any exceptions to the rule, and what organization(s) should be responsible for maintaining a repository of test reports and releasing the reports. If APSE E&V test results are not released, we must consider how the test results will be used and who will have access to the information.

It is recommended that E&V test results be released only when the APSE component(s) are being offered for commercial advantage. That is, results must be released if the component(s) are being sold or if the component(s) are/will be used on a government contract. Products which will only be used "in house" and products that are still in development may be exempted from public disclosure of

test results. (Developers should be encouraged to perform their own component testing until such time that their product is complete and ready for formal testing.)

It is also recommended that a repository of test results be maintained by the designated responsible DoD agency/office. Final reports should be sent to the National Technical Information Service (NTIS) to make the reports available to the public. The responsible E&V oversight facility should also maintain a historical database of evaluation and validation results. This database will provide a central information source and will be useful in assessing "trends" in APSEs over a period of time.

The E&V oversight facility should also publish an E&V consumer report on a semiannual basis. This report should include a list of tools/components that have been validated along with other pertinent data.

Recommendation: **Results of E&V testing of tools offered for commercial advantage should be maintained in a repository at the central E&V oversight facility and be made available to the public.**

2.5 Issue: Public Release of E&V criteria and tests

Should E&V criteria and tests be publicly released?

Discussion: It is recommended that E&V criteria and tests be made publicly available to users, developers, and managers. The suggested media would include the National Technical Information Service (NTIS) for documents, and ARPANET, MILNET, or tape for obtaining a complete set of tests. For tape copies, the requester would be required to provide blank tapes for the tests to be copied onto. E&V criteria and tests should be released by the designated responsible agency/office using the most recent version of the test set.

Release of the E&V criteria and tests will aid in ensuring that the evaluation criteria and individual test cases are valid since apparent errors will often be discovered by those whose products may be erroneously evaluated. Disclosure of E&V criteria and tests may also increase the objectivity of the evaluation by obtaining different viewpoints and interpretations of the standards and criteria.

Thus ambiguities may be identified and resolved. Developers will be more likely to accept evaluation results since they will have an opportunity to use the tests prior to formal E&V. Another advantage of releasing the tests is that new tests and improvements to existing tests may be suggested by the community. The motivation for developers to create/enhance tests is not only to demonstrate the capabilities of their own product but also to show the deficiencies that may exist in competitors' products.

Recommendation: **All E&V criteria and tests should be made available to the public.**

2.6 Issue: Definition of "CAIS Conformance"

Is "conformance to the CAIS" defined in a manner that will permit the development of a CAIS validation capability?

Discussion: CAIS conformance may be interpreted to be any of the following:

(1) To be CAIS conforming, an APSE must include all packages composing the CAIS and conform to each of them.

(2) Not all CAIS packages need be included in an APSE. An APSE must conform with only the included CAIS packages. An APSE including no CAIS packages is trivially conforming.

(3) Any other position intermediate between position (1) and position (2).

In the Draft Specification of the Common APSE Interface Set (CAIS), Version 1.0, it is stated that "Conformance of an implementation to the CAIS is established on a package by package basis." It is not specified whether all CAIS packages must be included to be considered CAIS conforming. The draft CAIS MIL-STD will become available in 1985 with the initial MIL-STD appearing in 1987. In order for the CAIS E&V effort to keep apace of the CAIS specification development, the policy concerning conformance should be explicitly stated for all aspects of the CAIS.

Recommendation: **The definition of CAIS conformance must be made explicit at the DoD level.**

2.7 Issue: DoD directive for CAIS

How will CAIS conformance of APSEs be encouraged/enforced?

Discussion: The establishment of a CAIS standard implies the intention that APSEs conform to this standard. Use of non-conforming APSEs on government contracts should be precluded by DoD directive. This directive should not be issued before all of the following have been accomplished:

- CAIS MIL-STD is established (expected 1987);
- CAIS validation suite is available;
- At least one conforming implementation is available.

Prior to this, a statement of intent to eventually issue such a directive is recommended. It will serve to alert APSE developers to a situation that will have great impact on their efforts, encourage early conformance to preliminary versions of the CAIS, and perhaps lead to greater interest and more rigorous review of the evolving CAIS specification.

Recommendation: **CAIS conformance of APSEs used on government contracts should be mandated by DoD directive. In the interim, the intention to do so should be made explicit.**

2.8 Issue: Encouraging use of E&V

How can the use of E&V technology be encouraged?

Discussion: Widespread acceptance of E&V will not happen by itself. The use of E&V technology must be encouraged for this vital function to occur. As E&V technology matures, the government should mandate the use of evaluation and validation where appropriate.

While E&V technology is developing, incentives should be established to encourage public use of E&V tools and techniques. These incentives could include, but are not limited to, the following. Director Defense Test and Evaluation (DDTE) could give credit in the Test and Evaluation Management Plan (TEMP) review for tools that have been evaluated and validated. Another incentive could be a

consumer report issued by the E&V central facility on a regular basis containing E&V results.

Recommendation: **Evaluation and validation should be mandated on all government contracts when E&V technology matures. Prior to E&V technology maturing, certain incentives should be provided to encourage public use of E&V tools and techniques.**

3.0 PROCEDURE ISSUES

This section presents recommendations related to procedure issues associated with the E&V Task. Procedure issues deal with performance of technical and administrative functions within the E&V Task.

3.1 Issue: Public review of E&V criteria and procedures

Should the E&V criteria and procedures be reviewed by the public prior to release?

Discussion: Unlike the Navy-led KAPSE Interface Team (KIT) which has its companion KAPSE Interface Team from Industry and Academia (KITIA) (which reviews and comments on documents produced by the KIT, and provides feedback and insight to the parent organization), the E&V Task has no such auxiliary group. Nevertheless, it is vital to the overall success and effectiveness of the E&V Task as a whole that the public sector be kept abreast of and contribute to this E&V activity.

Therefore, it is recommended that, in view of the fact that the need for this type of communication between the E&V Task (and its resulting technology) and the user community does exist, a mechanism should be provided which would promote and encourage industry participation in terms of periodic review and comment on the E&V technology.

While a need for public input has been identified, it is unclear at this point how to best implement this review process. A potential vehicle is a public review of E&V draft documents similar to the method employed by the KIT for public review of their draft Common APSE Interface Set (CAIS) document. Other options are open for consideration and discussion.

Recommendation: **The E&V Team should stage a public review of the E&V criteria and procedures prior to final release.**

3.2 Issue: Public coordination with other related efforts and technical groups

Should the E&V Team coordinate with the public during the E&V technology development process?

Discussion: The E&V Task should not only be concerned with addressing itself to that sector of the user community which is well versed in Ada and the Ada culture, and which regularly attends the various large Ada meetings and conferences (e.g., AdaTEC and AdaJUG), but should also consider the non-Ada groups and activities which might be potential users of E&V technology.

It is therefore recommended that, in order to promote this type of coordination and interaction, "road shows" be conducted to include briefings and/or presentations of papers on the E&V Task at various appropriate non-Ada technical groups, as well as input to publications such as IEEE Computer, Defense Electronics, ACM Communications, Government Computer News.

With respect to establishing coordination between the E&V Task and other related technical efforts, the E&V Technical Coordination Working Group (TECWG) is tasked with providing such coordination. Specifically, the TECWG is responsible for:

(a) performing a literature search for efforts related to the E&V Task; and

(b) developing a Technical Coordination Strategy Document which will identify related technical efforts, identify relationships between the E&V Task and each related effort, identify areas of mutual benefit, identify schedule impacts, identify the level of coordination required, and identify issues that require resolution to the mutual benefit of the tasks involved.

The E&V Task and other related efforts, as well as technical organizations, can mutually benefit from this exchange of ideas, lessons learned, etc.

Recommendation: **The E&V Team should coordinate with potential E&V users including non-Ada groups.**

3.3 Issue: E&V Technology Development Schedule

Should the E&V technology development schedule be related to software tool availability?

Discussion: One approach is to develop the E&V technology independent of the sequence in which the community actually undertakes tool development. This could possibly result in tools being available with little E&V technology with which to assess tool capability -- a situation that flourishes today with many unacceptable consequences.

An alternate approach is to coordinate E&V technology development with actual emergence of the tools requiring evaluation and validation. To some extent this is already being performed by the E&V Team in the instance of the Common APSE Interface Set (CAIS). The development of the validation suite closely mirrors the evolution of the CAIS. This technique will provide a validation capability for the CAIS in a timely manner. A similar approach appears advisable for other tools so that as they become available the associated E&V criteria and test suites are also available to assist the public in tool selection.

Recommendation: **Particular tool E&V technology development should closely follow tool availability.**

3.4 Issue: Weighting of test results

Should E&V tests be weighted?

Discussion: An E&V measure based solely on the percentage of tests passed may be a misleading indicator of the usefulness and correctness of the APSE components. To overcome such a problem the tests must be organized into a structure that is based on the functional requirements of the components under test. In turn, these functional requirements may be given weights so that demonstrating compliance with a particular requirement is "worth more" than compliance with a requirement of lesser importance. Weighting of tests is particularly important in cases where policy does not dictate that all validation tests must be passed. Many users believe that a relatively high score for conformance is good enough. When significant areas of conformance are not met, weighting should be used to make clear to the community the seriousness of the deficiency.

Weighting of tests is not as applicable to evaluation tests. The importance of particular metrics should not be dictated to the user community. Our goal is not to achieve a "goodness" value but to indicate that some tests may have more widespread implications. Weighting may be applicable to an evaluation factor where many subfactors are evaluated to achieve a composite value. For example, many subfactors will be considered in evaluating the portability of a tool. Subfactors may include tool interfaces (many tests), tool size, structure, and the language in which the tool is written. Even if a tool was not written in Ada, it could score very high in portability based on high scores in the other subfactors. This is one example where weighting of tests may be appropriate to evaluation.

Recommendation: **Test results should be weighted so that the scope of an individual test may be more easily understood.**

3.5 Issue: Reader's Guide for interpretation of test results

Should a guide for test results interpretation accompany the results?

Discussion: Because of the importance of sharing the E&V results with managers and organizations just becoming involved with Ada, in addition to application analysts and programmers, an E&V Test Report Reader's Guide should be developed to aid in understanding and interpreting the E&V Test Report. The Reader's Guide should provide guidance as to the importance of particular evaluation factors and performance characteristics relevant to particular applications. For instance, I/O would be very important to Command, Control, Communications, and Intelligence (C^3I) applications, where databases are heavily used. The Reader's Guide might give a range for the number of I/O operations per second that would be considered acceptable for such a system. For avionics and armament applications, one important aspect might be the mapping of interrupts to task entries.

The Test Report Reader's Guide would also discuss the relationship among the metrics (e.g., the greatest degree of portability may only be achieved by the sacrifice of efficiency and vice versa). This would help the APSE procurer in making trade-off decisions. Guidance would also be given regarding the applicability and importance of specific criteria to each life-cycle phase. The Guide would only provide guidance; its intent is not to say whether an APSE is good or bad, but to

provide the necessary information for a program manager to draw his own conclusions and make a decision based on the needs of his particular program. A copy of the Test Report Reader's Guide should be distributed along with the test results as a companion document.

Recommendation: **E&V test results should be accompanied by a reader's guide aimed at helping non-Ada, non-software-oriented program managers understand and better use the information.**

3.6 Issue: Application-specific benchmarks

Should application-specific benchmarks be developed to aid in tool evaluation?

Discussion: Performance characteristics of APSE tools and Ada programs generated from said tools are an important factor in evaluation. Performance is measured relative to some set of trial cases or benchmarks. Performance for a particular application can best be evaluated with a benchmark from that application area.

Such application-specific benchmarks should be developed and collected as part of the E&V Task. Contribution of benchmark cases should be solicited from projects in the areas of avionics, armament, C3I, space, etc. In applications areas where benchmark cases cannot be collected, they should be developed by the E&V Task.

Recommendation: **Application-specific benchmarks should be collected and developed as part of the E&V Task.**

3.7 Issue: APSE definition

How should APSEs be viewed for E&V assessment (e.g., as a whole, as a sum of the parts)?

Discussion: STONEMAN provides only suggestions for the tools to be included in an APSE and leaves even this open-ended. An enumeration of expected tools or a detailed taxonomy of APSE tools should be developed, and non-tool-oriented views of the APSE explored.

The tools used in an APSE may be chosen to support a particular application, architecture, methodology, or combinations of these factors. The configuration of a tool-set may be quite varied. The issue arises of evaluating and validating tools and tool-sets in various configurations. Are all members of tool-sets to be re-E&V'd together for each configuration, or may new tools be added after individual E&V? Restated, are APSEs to be considered for E&V as wholes, or as the sum of the parts?

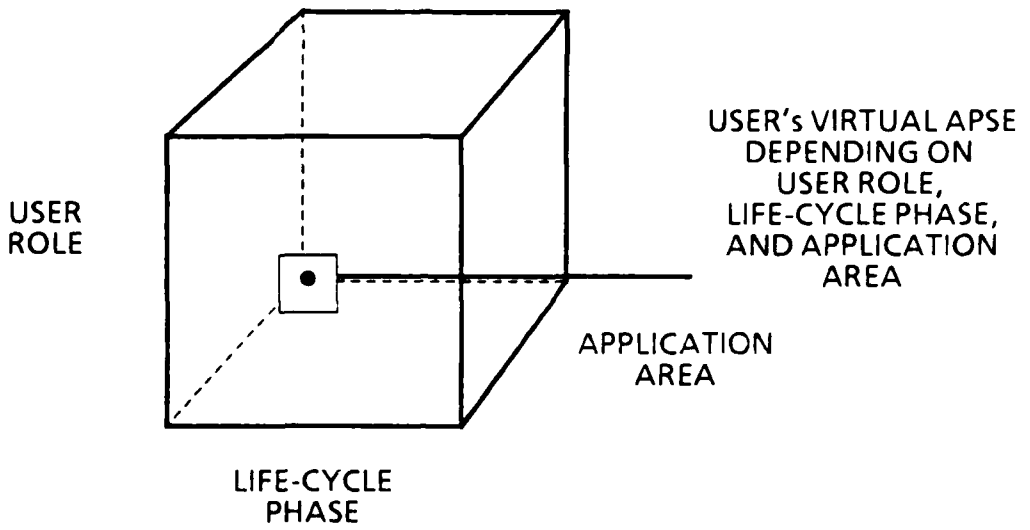
A compromise is to define a set of core tools, perhaps equivalent to the MAPSE, which would contain capabilities common to the broad range of applications. Supplementary tools, which provide capabilities needed for a particular application, methodology, etc., could be E&V'd individually, and with the core tool-set as an entity. The issue is confused by the potential need for multiple core tool-sets to provide capability to different architectures or methodologies.

If an APSE is a related collection of Ada life-cycle support tools, then user-view APSEs may have to be described. The user's view of the APSE may be dynamic; changing with the project phase, user role, or application area. The cubic model is helpful.

The application area determines the plane of tools effective for the application. The tools available to the user at a particular time are determined by the user's role in the project and the life-cycle phase under way. Other factors may affect the user's virtual view of the APSE, including the methodology in use.

Alternate views of the APSE may serve to suggest criteria for E&V assessments. These views may include a project database perception of APSE functions, or a process executive view in which function availability is limited by an underlying APSE executive.

<u>Recommendation</u>: An enumeration of expected tools or detailed taxonomy of APSE tools should be developed and non-tool-oriented views of the APSE explored.



4.0 STANDARDIZATION ISSUES

This section presents recommendations related to standardization issues associated with the E&V Task.

4.1 Issue: DIANA

Should DIANA become a standard?

Discussion: There are many APSE tools that need to operate on a "digested" form of Ada programs. Such a digested form contains the Ada program described by its syntactic and semantic content rather than its simple source form. To require each of those tools to produce such a form before using it promotes duplication of functionality and propagation of complexity. There should be only one tool to digest Ada source programs, and a standard, compact digested form that all APSE tools may access.

There is such a form currently in widespread use. It is called DIANA and has been carefully described at the appropriate level in the DIANA Reference Manual. It is not yet a standard. Therefore, most users corrupt it in some way as they

implement it. The proliferation of DIANA dialects directly prohibits the use of DIANA as the form accepted by all APSE tools, especially when those tools reside on different APSEs.

Recommendation: **DIANA should be standardized as soon as possible and should be the only digested form of Ada progams used by APSE tools.**

4.2 Issue: APSE terminology and notation

Is there a need for industry-wide standardization of APSE terminology and notation?.

Discussion: When vendors describe APSE tools, they frequently use different names to refer to functionally equivalent tools, or use the same name to refer to functionally different capabilities. A "configuration management tool" is an example of a term commonly used to refer to anything from version control to discrepancy reporting tools. The only resolution to this problem is standardization of terminology.

Standardization may be introduced basically in two ways:

(1) Industry can develop a standard,

(2) The E&V effort can impose a standard.

Recommendation: **Precisely defined standard terminology and notation must be established in the near term. The preferred method is to encourage industry to develop its own standard.**

5.0 RELATED ISSUES

Several issues of importance to the evaluation and validation of APSEs could not be assigned to the previous categories and are collected here. These include issues whose technical, policy, or procedural impact is not fully understood by the Recommendations Working Group. They are generally open questions needing further investigation or assessment.

5.1 Issue: APSE Maintenance Tools

What types of APSE maintenance tools are needed?

Discussion: The E&V process must encourage the development of APSE tools which aid program maintenance activities by producing maintainable code, by providing support to a new programmer in understanding a program to be maintained, or by other means. It is recognized that program repair (bug fixing) and program enhancement are two different activities considered to be maintenance, and that these activities use quite different approaches. Both activities require tools with extended capabilities beyond traditional development or debugging aids, especially the use of existing information to isolate program defects or aid the review and incorporation of new requirements and specifications with existing ones. System-level maintenance aids to identify maintenance trends or problem areas are to be encouraged.

Recommendation: **Technology which supports the repair or enhancement of programs must be further developed. New tools must be constructed because traditional development aids are inadequate for post-deployment program maintenance.**

5.2 Research Issues

The state of practice regarding evaluation and validation of tools individuall, and collectively is still at an early and tentative stage. In generating this set of recommendations for APSE E&V, many issues arose which were poorly understood or were based on unimplemented technology. Resolution of these issues is not practicable until more background information is developed through research. Research topics important to APSE E&V are presented here.

5.2.1 Metrics Definition

Validation and evaluation metrics are needed. Measurement of software tools is rudimentary, with compilers having received the most attention. Not all compiler measures can be extended meaningfully to other tools; many new techniques need to be developed. Some include quantified measures for human-tool interaction such as adaptability, usability, or ease of learning, or quantified measures for configuration management functions.

A distinction between E&V of tools and of whole APSEs is apparent. Evaluation of the usability, efficiency, or adaptability of an environment is different than for an individual tool. APSE-level metrics and procedures must be developed beyond the mere summing of tool scores.

5.2.2 Revalidation Triggers

Tools and APSEs should be revalidated after major revisions. However, the definition of "major revision" must be determined. The percentage of lines of code changed, added, or deleted may contribute to this definition. The re-hosting of a tool or APSE may be a trigger for re-E&V, but for computer hardware families and operating system families the boundary becomes vague. The time period for periodic re-E&V should be based on some nominal project length.

5.2.3 Inter-Tool Communication

Data structures should be standardized for communication between tools in an environment. Current approaches, such as UNIX, use the lowest common denominator approach and pass only byte streams. Higher-level interface objects, such as DIANA, might be possible and desirable.

5.2.4 Language Feature-specific Benchmarks

The performance requirements of mission-critical software systems cover a wide range of demands on APSE capabilities. Different missions place different but quite specific and identifiable demands on features of APSE capabilities. Demands on the Ada compiler range from the speed requirements of an avionics or armament system to the data transfer and communications needed for C3I systems. The avionics or armament applications may depend on the task interrupt entry reaction time for its success while the C3I system may depend on the efficiency of a low-level I/O package implementation for a particular device. These can be termed application-specific performance requirements. Benchmarks that allow evaluation of these and similar features are needed. Both may be concerned about task activation, termination, or rendezvous overhead. These would be common performance requirements. A suite of common benchmarks for compiler evaluation is needed. Other APSE tools may require analogous benchmarks to measure performance.

5.2.5 Independent Evaluation of E&V Technology

Concurrent with any development of technology in support of software engineering should be an independent development of ways to measure its validity and evaluate its performance.

5.3 Issue: Methodology Biases in the Evaluation Process?

How can methodology biases be minimized during the evaluation process?

Discussion: Many APSE components undergoing evaluation support particular software development methodologies. A quality assessment of a component might, therefore, be construed to be a comment on the quality of the methodology. The evaluation of software development methodologies is beyond the current state of the technology. Existing methodologies appear to be more supportive of particular application areas. No methodology is generally suited for all application areas. In any event, methodology assessment is beyond the scope of the E&V effort, which is intended to evaluate APSE components; no implication to the contrary should be made in reporting E&V results.

Software quality metrics are based on assumptions concerning what constitutes "good" software development practices. Such practices usually consist of traditional, top-down structured design techniques which support the traditional life-cycle model. Those APSE components, therefore, that support this class of methodologies would be expected to be judged high in quality. Components that support alternative methodologies may be evaluated lower in quality because they involve different life-cycle models or different assumptions concerning "goodness". Evaluation results would, therefore, be systematically biased in favor of tools supporting certain methodologies.

Innovative software practices may introduce radically different software development models. An example is the functional life-cycle model. Existing quality metrics may not be able to fairly and adequately assess the quality of tools supporting the functional lifecycle. As new technologies are introduced, the evaluation suite must be extended. Otherwise, evaluation results will always be biased towards older technology.

Not all metrics are applicable to all components. Care should be taken not to apply inappropriate metrics lest results be misinterpreted and reflect on the quality of the tool or the methodology it supports.

Recommendation:

- **No endorsement of a particular methodology should be intended.**
- **Evaluation results should not be biased in favor of certain methodologies.**
- **Metrics that are not applicable should not be reported.**
- **The evaluation suite must be extensible. As new technology is introduced, new evaluation metrics must be included.**

5.4 Issue: Legal Considerations Relating to E&V

What are the legal implications regarding warranties or liabilities associated with E&V'd software?

Discussion: The legal implications of an E&V'd APSE component are unknown with respect to liabilities if the performance of the component fails to match its

described characteristics. The relationship of protections offered by copyrights or patents on software products to the E&V process must be investigated. Proprietary source code and/or trade secrets must be protected during the E&V process.

Recommendation: **Legal implications require further investigation.**

5.5 Issue: E&V of Mixed-Language APSEs

Should tools containing components not written in Ada be acceptable in APSEs?

Discussion: As APSEs are initially being developed, it will be desirable to augment them with existing tools. It is unlikely that these existing tools will be written in Ada. To preclude use of such tools would cause either delayed introduction of APSEs while existing tools are translated to Ada or introduction of APSEs with insufficient power or range.

On the other hand, allowing APSE tools to be written in a language other than Ada means that the portability and maintainability of those tools are questionable.

Recommendation: **Non-Ada or hybrid tools should be acceptable in early APSE tool-sets as long as the intent is to eventually migrate toward all-Ada tools. E&V practices should acknowledge the existence of such non-Ada tools and be able to evaluate and validate them. That is not to say that a non-Ada tool should score the same on evaluation as an identical all-Ada tool. For example, an all-Ada tool would probably score higher on maintainability than a non-Ada tool. However, it should be unacceptable to reject a tool from E&V processing simply because it is not written entirely in Ada.**

This recommendation should not be construed as encouraging or even allowing the development of new tools for APSEs in anything except Ada.

After a suitably long grace period, this allowance should be discontinued.

5.6 Issue: E&V of Distributed APSE functions

What is the nature of E&V for multiple versions of an APSE function hosted on a variety of hardware configurations?

Discussion: The APSE described by STONEMAN implicitly assumed a central minicomputer with attached terminals as development host. Current hardware technology trends are toward distributed environments, including networks and independent workstations. APSE capabilities previously resident on a single computer may now be moved to or spread over several nodes in a system. For example, text editing and formatting may be done on an intelligent workstation, freeing a larger central computer from these tasks. Communications capabilities, such as electronic mail, are important APSE functions in distributed environments, and standard protocols may be required. Adequate and efficient allocation of functions to nodes in a distributed APSE may be required for APSE E&V.

Independent workstations may be employed in program development without interconnections. Guidelines for transfer of data (such as physical media and formatting standards) may be reviewed by the E&V process.

Recommendation: **Distributed environments present significant new problems for E&V of APSE functions spread among several nodes in the environment. Research is needed to provide the technology base for E&V of distributed APSEs. E&V criteria should encourage APSE functions which may be easily adapted to a variety of APSE hardware support architectures.**

5.7 Issue: Run-time support standardization

Should run-time support systems be standardized?

Discussion: To support portability of applications and tools, run-time support standards beyond those provided by CAIS need to be investigated. Some standards or guidelines for real-time executives, run-time libraries, and other implementation-dependent features for classes of applications or families of processors may be possible.

Run-time support standards for service-specific computers will have to be developed by the services, and are likely to be somewhat application-oriented. An important consideration will be run-time compatibility with existing applications written in other languages.

Run-time support for commercial computers may be considered from either host or target views. Run-time support in a host-type system depends on coexistence with existing operating systems and thus will be widely varied.

Run-time support of Ada tasks is a critical performance issue for target computers. Some basic models exist, but it is far from certain that these represent optimum approaches. Continued investigation of run-time support of tasking is needed before standards can be recommended.

E&V of run-time support, whether standard or non-standard, is currently an ad hoc approach, depending largely on benchmarks and performance evaluation measures. Considerable effort to gain understanding of E&V of run-time support systems is needed.

Recommendation: **Run-time support standardization should be investigated.**

6.0 REFERENCES

Buxton, J.N., "Requirements for Ada Programming Support Environments -- 'STONEMAN'," U.S. Department of Defense, Februa y 1980.

Castor, Virginia L., "Evaluation and Validation (E&V) Plan, Version 1.0," AFWAL/ AAAF, Wright-Patterson AFB, OH 45433, 30 November 1983.

KIT/KITIA CAIS Working Group, "Draft Specifications of the Common APSE Interface Set (CAIS), Version 1.1," Ada Joint Program Office, 30 September 1983.

(LRM): "Reference Manual for the Ada Programming Language, ANSI/MIL-STD-1815A," U.S. Department of Defense, January 1983.

APPENDIX C

E&V REQUIREMENTS WORKING GROUP REPORT

APSE
EVALUATION AND VALIDATION (E&V) WORKSHOP
AIRLIE, VIRGINIA

April 2-6, 1984

APPENDIX C

E&V REQUIREMENTS WORKING GROUP REPORT

TABLE OF CONTENTS

I. E&V REQUIREMENTS WORKING GROUP OVERVIEW C-3

1.0 List of Members C-3
2.0 Goal C-3
3.0 Accomplishments C-3

II. E&V REQUIREMENTS C-6

1.0 RECOMMENDED REQUIREMENTS ON THE E&V TEAM C-6
1.1 Phased Development of an APSE E&V Capability C-6
1.1.1 Introduction C-6
1.1.2 Background C-6
1.1.3 Effect on E&V Requirements C-7
1.1.4 Recommendations C-8
1.2 E&V Product Quality Guidance C-9
1.2.1 E&V Team Quality Program C-10
1.2.2 Quality Control of E&V Products C-11

2.0 RECOMMENDED APSE EVALUATION REQUIREMENTS C-12
2.1 Requirements Relating to Methodology Supported by an APSE C-12
2.1.1 Background C-12
2.1.2 Motivation C-13
2.1.3 Discussion C-15
2.1.4 Position C-15
2.1.5 Reference Manual Recommendations C-15
2.1.6 Sources C-17
2.2 Requirements Relating to Life-Cycle Support by an APSE C-17
2.2.1 Background C-17
2.3 Requirements Relating to Application-specific Concerns C-18
2.3.1 Background C-18
2.3.2 Motivation C-19
2.3.3 Example Application Areas C-21
2.3.4 Issues C-22
2.3.5 Conclusions C-22
2.4 Requirements Relating to Evaluation of Simulation Support and Embedded Software Development C-23
2.4.1 Background C-23
2.4.2 Motivation C-23
2.4.3 Examples of Desired Tools/Features C-23
2.5 Requirements for the Evaluation of Inter-Tool Interfaces C-24
2.5.1 Background C-24
2.5.2 Motivation C-24

3.0 GENERAL CONCERNS C-27

4.0 REQUIREMENTS DOCUMENT OUTLINE C-29
4.1 General Introduction C-29
4.2 Approach C-29
4.3 Product Quality Guidance C-29
4.4 APSE E&V Requirements C-30
4.5 Summary/Conclusions C-30

I. E&V REQUIREMENTS WORKING GROUP OVERVIEW

1.0 REQUIREMENTS WORKING GROUP MEMERS

Chairperson--Timothy E. Lindquist, Virginia Tech.

Members:

Maj Daniel Burton, ESD
Bard Crawford, TASC
Mr. Nelson Estes, ASD
Kathleen Gilroy, Harris Corporation
Asha Kant, Litton Applied Technology
Michael Meirink, Sperry Corporation
James Parlier, General Dynamics
Helen Romanowsky, Rockwell
Andres Rudmik, GTE Networks
Paul Scheffer, Martin Marietta Denver Aerospace.

The positions presented in this report were formulated by the above members and have benefited greatly from review and discussions with the government members of the working group.

2.0 GOAL

The goal was to formulate industry input to the E&V Team in the form of specific technical topics that need to be addressed by the Requirements Document.

3.0 ACCOMPLISHMENTS

The E&V Workshop Requirements Working Group accepted three activities relating the industrial position papers and E&V Requirements Document.

The first activity undertaken by the group was to review the position papers of the industry participants and the requirements implied by those positions. Discussions focused on the following topics:

(1) Life-Cycle and Methodology Support;

(2) Requirements Based on Application Concerns;

(3) Evaluating Inter-Tool Interface Complexity.

Each topic is represented by sample requirements and a rationale in Section II.

The second activity undertaken was to examine the relationships between the Requirements Document, Reference Manual, and Guidebook. In a joint meeting with the Reference Manual Working Group, sample Requirements, Reference, and Guidebook entries were formulated and discussed.

The third activity of the working group was to address the draft outline for the Requirements Document that is contained in Attachment 1 of this report. The group addressed and formulated suggested contents for two subsections of the document, and discussed the framework that should be used to present the specific requirements for evaluating an APSE. Section II, 1.1 of this report addresses a phased development of E&V technology in the form of a suggested framework for the E&V Team. An approach is discussed that allows orderly and incremental increases in E&V technology while addressing the immediate and long-term APSE evaluation needs. The suggestion deals with the reality that:

Today we have "non-APSEs",
Tomorrow we will have APSEs, and
In the future we will have "SUPER APSEs".

The phased development approach to E&V would incrementally provide evaluation technology appropriate to APSEs as they evolve. This approach belongs in Section 1 (Goals) of the E&V Requirements Document outlined in Attachment 1.

Section II, 1.2 of this report addresses the need for E&V product quality guidance. The working group's proposal recommends that all products of the E&V Team be subjected to configuration management and quality assurance. For our purposes, products included the documents generated by the team and any

evaluation or validation test or procedure initiated by the team. Sections II ,2.0 and II, 3.0 address required evaluations that need to be performed on APSE components.

II. E&V REQUIREMENTS

1.0 RECOMMENDED REQUIREMENTS ON THE E&V TEAM

1.1 Phased Development of an APSE E&V Capability (Near- to Far-Term Approach)

1.1.1 Introduction

The overall objective of the E&V Task is the development of a technology which can be applied to the Evaluation and Validation of APSEs. This section addresses the approach that should be taken in the development of this technology and its effect on the stated requirements for the E&V technical effort. It is based on the E&V Plan, the position papers presented at the E&V Workshop, and related discussions by the Workshop participants.

1.1.2 Background

Currently available or proposed APSEs implement a variety of conceptual models, not all of which conform to the STONEMAN concept of an APSE. Also, currently available Ada programming support consists of more limited tool-sets (e.g., compiler in a non-Ada/non APSE environment).

The trend in current software development practices is toward emphasis on structured specifications promoting testability, application of testing earlier in the life cycle , and automation of the testing process.

The users/buyers/developers of APSEs have immediate requirements for the E&V of these initial APSEs and tools/tool-sets. The NBS taxonomy includes a prioritization of some of these needs, and indicates the following list of APSE features as having the highest priority:

- compilation (transformation)
- editing (transformation)
- formatting (transformation)
- configuration control (management)
- Ada library management (management)

- on-line command and error assistance (input/output)
- type analysis (static analysis)
- interface analysis (static analysis)
- cross reference (static analysis)
- tracing/debugging (dynamic analysis)
- optimization (transformation).

Other immediate requirements are:

- file administration/database management
- language features/RTE
- loader
- commands/command procedures/parameters
- documentation/maintenance support
- cost information.

It is perceived that APSEs will evolve to the extent that the current concept of an APSE tool, such as a compiler, will not be an appropriate subject for E&V. Rather, emphasis will be on functionality, interfaces, and data at both microscopic and macroscopic levels.

1.1.3 Effect on E&V Requirements

The immaturity of the current state of the art in Software Support Environments and E&V technology warrants a phased approach to the development of an APSE E&V technology. The immediate needs of the Ada community in advance of this technology demand a phased approach. Long-term needs should not be ignored, with near-term efforts incorporated into the far-term technology, and guiding the far-term efforts.

The E&V Team should expect and plan for changes in requirements due to the evolving nature of:

- APSE concepts/composition/use
- CAIS specification and other standards
- E&V practices/techniques/methodology
- user needs.

The impact of an Ada components industry may also force reassessment of the initial E&V requirements.

1.1.4 Recommendations

(a) Near Term (First Phase)

It is recommended for the near term that the definition of E&V requirements be such that it minimally provides:

- support for E&V of those individual APSE features needed immediately;
- encouragement for standardization (formal or de facto) of common interface classes (DIANA, textfiles, RTE, program library, CAIS, etc.);
- support for E&V of APSE functionality in a "real-world" environment;
- initial global E&V of a baseline APSE;
- well-specified requirements;
- an initial taxonomy of an APSE.

(b) Intermediate Phases

Once an initial E&V capability has been provided, emphasis can shift toward developing expanded E&V technology. E&V capabilities/activities at these intermediate phases could include:

- evaluation of functional components;
- evaluation of protocols used by the functional components;
- development of the CAIS validation capability (CVC);
- definition of semantics and notation for an APSE;
- definition of "CAIS conformance";
- evaluation of additional functionality, such as simulation;
- development of new metrics for evaluation, especially for the "ilities";
- refine, replace, remove requirements of previous phases;
- evaluation of support for Ada-based PDLs/RSLs;

- increased emphasis on user or operational perspective and human factors;
- increased emphasis on host/target relationship;
- increased emphasis on global issues.

(c) Far Term (Phase X and beyond)

Some participants at the workshop felt that the long-term goal is a generalized E&V capability, which will be useful independent of programming language. Other far-term requirements/capabilities include:

- capability to E&V a variety of project-specific, application-specific, or life-cycle/methodology-specific APSEs (including distributed APSEs and APSEs supporting embedded systems applications);
- automated E&V support and use of E&V early in the APSE life cycle;
- incorporation of new standards/functionality;
- refine/replace/revise/remove previous phase requirements;
- E&V technology carries into applications-level programs.

1.2 E&V Product Quality Guidance

Product quality guidance is provided to establish a framework for the identification, definition, allocation, execution, and reporting of quality-related activities for the E&V Task. A product quality program shall be established for the E&V Team products. This program will contain elements of a managerial and technical nature, whereby E&V tools may be subjected to formal acceptance criteria by the user and, when contracted, by the cognizant customer agency. In order to ensure proper management control, communications, and visibility, a life-cycle approach should be employed to optimally effect a time-phased, project-oriented product quality program. This will also allow for an effective means of establishing cost, schedule, and performance parameters conducive to good management practices. To this end, the provisions of MIL-STD-SDS, MIL-STD-1679, 483 and 490, and MIL-S-52779 are applicable as guidelines and references.

1.2.1 E&V Team Quality Program

The E&V Team product quality program shall consist of the following disciplines, as a minimum:

Establishment of a product group within the E&V team.

The functions of the organization shall be:

- Establishment of test programs that shall have the following characteristics:
  - -- self-contained;
  - -- independent;
  - -- can be grouped to run as a module;
  - -- do not require factoring to use;
  - -- self-checking where possible;
  - -- documented as to purpose, inputs, outputs, and results;
  - -- are generally applicable;
  - -- an adequate number of test levels to encourage structured testing;
  - -- stress and off-nominal test case use;
  - -- authorize/require use of test notebooks.
- Checklists (quality and completeness of requirements audits shall be used. Results shall be reported to E&V Team management and other appropriate agencies, if necessary).
- Questionnaires determining whether:
  - -- a particular standard or requirement is applicable and, if applicable, how adequately do E&V products comply with the requirements or standards?
- Configuration management:
  - -- identification, control, and status accounting practices shall be established and maintained;
  - -- baselines shall be established and based on formal reviews and audits conducted against the products of each phase of the life cycle of APSE formulation;
  - -- changes to APSE E&V configuration and/or tools will require the approval of an in-place board chartered with the responsibility of change tracking and

implementation;

- Version Description Documents (VDDs) (Ref. MIL-STD-483) shall be produced that identify approved changes and that incorporate problem report information that resulted in a change from the previously approved baseline;
- procedures for reporting problems shall be established that include: a numbering system, problem priority, category, corrective action/approval status, and guidelines for problem descriptions;
- status accounting on all baselined products shall be accomplished.

Use of automated configuration management tools is encouraged.

1.2.2 Quality Control of E&V Products

The products of the E&V Task shall be subject to quality control to provide conformity and consistency. The quality control process should entail a documentation and code review. The actual product, i.e., the code, shall be compared against the desired product, i.e., the requirements and the design. A traceability analysis shall be performed to detail the relationship between E&V tools and E&V requirements. A resulting traceability matrix should be produced and made an integral part of the system documentation.

A more qualitative review of the documentation (development and test) shall be conducted. Comments should fall into three general categories:

(1) violation of an applicable standard,

(2) technically complete and consistent, not subject to misinterpretation, and does not violate previous specifications, and

(3) typographical and style considerations.

2.0 RECOMMENDED APSE EVALUATION REQUIREMENTS

2.1 Requirements Relating to Methodology Support by an APSE

(1) Assess to what degree an APSE supports specific development methodologies;
(2) Assess the degree of flexibility and extensibility of an APSE to support methodologies;
(3) Assess how an APSE's tools and supported methodologies improve productivity;
(4) Assess the ease of transition among (life-cycle) phases.

2.1.1 Background

A methodology is a body of methods, rules, and postulates employed by a discipline.

Methodologies vary in at least the following key ways (when, by whom, on what):

- The problem domain or environment of the methodology.
- The information base the methodology must address.
- The set of audience skills in using the methodology.
- The outcomes to be achieved by using the methodology.
- Ease of (publicly) evaluating the methodology (how well articulated it is).

Using these variables, the following methodology types are useful for discussing the suitability of an APSE to a methodology:

(1) Ad hoc -- This is a trial and error approach. Methodology is not planned in advance, but is based upon immediately prior results.
(2) Artistic -- This approach relies upon the skill of the master. Broad tasks can be described, but actual execution is dependent on the skill level of the staff.
(3) Heuristic -- The major ways of approaching a problem and a variety of specific techniques are described in writing. Closing criteria, if present, are at least partly arbitrary.

(4) Procedural -- The process is explicitly articulated. It identifies major steps (perhaps substeps), the information handled at each step, key measures and metrics, and completion and quality criteria. The intent is to influence the order in which decisions are made and to improve visibility and control over the process.

(5) Formal -- Notations used by the methodology are machine-analyzable. That is, the syntax of the language can be formally defined. Tools can support a degree of completeness or consistency checking. Often they are usable by a wide range of specific methodologies.

(6) Rigorous -- Notations used by the methodology have precise semantics. The notation has sufficient operational definition to allow empirical validation of constructs and their representations. (Examples are Petri Nets, Finite State Machine.) This includes any representation that can be executed. Tools supporting specific rigorous methodologies are often not usable by other specific methodologies.

From a user's viewpoint, a software development methodology has these key components:

Notations	--	language used to describe the system to be developed.
Methods	--	techniques to develop and to determine the development's compliance to criteria.
Tools	--	automated support for handling notations and for encouraging/ enforcing methods.
Procedures	--	written description of the proper use of notations, methods, and tools.

2.1.2 Motivation

There are several motivations for raising this issue:

(1) STONEMAN expects to support methodologies: "Ada Program Support Environments (APSEs) which are constructed by extensions of the MAPSE to provide fuller support of particular applications or methodologies." Concern for the disciplined use of the APSE motivated the DoD METHODMAN effort.

(2) Organizations will strive to improve visibility and control over the

software development process. That is, the entire process must not be totally ad hoc or artistic. The process must be appropriate to the task. Each of the methodology types has an appropriate use such as innovation (artistry must not be stifled) or mechanization (procedural, formal, rigorous). Productivity gains can be achieved by improving tool capability and increasing tool use.

(3) Interest in the front end of the software life cycle is high. The majority of errors occur then; the later they are detected, the more costly they are to repair. Many have developed or are developing tools to support an "Ada Program Design Language". A major tool-set under the auspices of the Joint Services Software Engineering Environment (JSSEE) Committee is the Distributed Computing Design System (DCDS). The Army CECOM has contracted the development of methodologies whose notation for requirements specification and design is pure Ada (Ada Formulation Methods Study). These methodologies promote the notion of merging the process of conceiving (creating) the design with the process of recording it. The latter two are examples of rigorous approaches. A key issue which emerges is what information beyond Ada is needed to describe design? Today, contractors must look beyond tools written in Ada or conforming to a CAIS interface.

(4) APSEs that claim to support one (or more) independently available development methodologies may do so in one of two ways: first, by providing tools of a general nature which are so parameterized and interdependent that a regimen of their use realizes a specific methodological scheme (e.g., design graphics + text + report generator to yield Yourdon DFDs, mini-specs, and data dictionary). The regimen of use to achieve this methodology must be included with the APSE to support the claim. Second, an APSE may include a tool-set which specifically implements a development methodology, with no variation on its regimen of use. This case may be exemplified by future APSEs which include SREM, SADT, or PSL/PSA, for example.

2.1.3 Discussion

From a buyer's viewpoint, the key questions are:

(1) Can this APSE support my methodology?

(2) What tools are available to support a methodology?

(3) Is this methodology and its supporting tool-set applicable to my problem (or project)?

(4) What are the workproducts of the methodology?

Tools may be integrated with respect to other tools (i.e., program-callable interfaces to the database or inter-tool data), with respect to the user interface, or with respect to a methodology. Thus, tools used in the context of a methodology may form an integrated set regardless of other aspects of integration.

2.1.4 Position

The crucial question is "What is needed to evaluate the suitability of an APSE to one or more methodologies?" APSE support for methodologies cannot be meaningfully evaluated by microscopic analysis of tools and their interfaces. There are, however, several reasonable approaches which can be taken:

(1) Identify information items and questions which could serve as a basis to begin evaluation;

(2) Develop benchmark problems for industry or academia to use to exercise methodologies and their tool-sets;

(3) Issue guidelines for conducting data-gathering exercises. The guidelines should address controls, transition between tools and methodology tasks, errors, tool and people performance, and analysis of results.

2.1.5 Reference Manual Recommendations

The following strategies and criteria are recommended for evaluation:

(1) A tool (or tool-set) said to support a methodology should supply the following information:

(a) Description of the methods suggested via an explanation

of the analysis techniques, underlying formal foundation, and rationale.

(b) Definition of the notation supported --
  (i) Formal Definition (syntax and semantics)
  (ii) Cite and explain Ada compatibility
    a. Notation is a subset of Ada -- identify exclusions;
    b. Notation is a subset of Ada plus extensions -- identify exclusions and extensions;
    c. Notation is a superset of Ada -- identify extensions;
    d. Notation is not Ada-based -- if applicable, explain mapping to Ada.

(c) Describe tool-set capability -- inputs, outputs, functions, manuals, etc.

(d) Provide procedures (layer by detail) --
  (i) Overview;
  (ii) Heuristic (major process steps);
  (iii) Procedural.

(2) To evaluate the productivity of using an APSE environment according to some methodology, E&V should consider:
  (a) Methodology and APSE support for controlling the development;
  (b) Definition of errors;
  (c) Size and complexity of the application problem;
  (d) Level of training;
  (e) Data-gathering and analysis techniques,
  (f) Number of participants.

The METHODMAN guidance for conducting experiments can serve as a useful source for producing guidelines.

(3) To evaluate the extensibility and flexibility of an APSE in supporting multiple methodologies, E&V should consider the following for multiple projects or methodologies:

(a) How easily is the notation tuned?
(b) How easily can the decision rules encouraged by the tool-set be tuned?
(c) How easily are the procedures altered?

(4) To assess the ease of transitioning across tasks or phases, E&V should consider:
(a) Information passed (to next phase);
(b) Information saved (as history, hidden);
(c) Information sunk (no longer needed);
(d) Information lost (should have been passed).
(e) Activities (actions) interacting among phases.

2.1.6 Sources

- METHODMAN.
- "Ada Design Language Concerns," Grau & Comer, Ada Technology Conference, March 27-28, 1984, U.S. Army, Ft. Monmouth.
- "A Software Engineering Environment for the Navy," NAVMAT SEEWG.

2.2 Requirements Relating to Life-Cycle Support by an APSE

(1) A classification of APSE tools shall be constructed to indicate which life-cycle phase is addressed by each tool.
(2) APSE evaluation shall include the degree to which each tool supports the life-cycle phase it addresses.

2.2.1 Background

To simplify the APSE concept-of-use, its features should readily relate to users' needs within the framework of the software development life cycle. The criteria for making this assessment consist of relating each tool to the life-cycle phase it supports, and, for all tool-sets, assessing the degree of comprehensive coverage of the life cycle. Tools that do not apply to a specific phase, and so support multiple activities in a utility sense (editors, report generators, etc.), should be so identified.

2.3 Requirements Relating to Application-specific Concerns

(1) Evaluate based on tool-set capability and the efficiency of their use with respect to supporting application-specific tasks.

(2) Application-specific APSE evaluation requirements include:

- Capability of the APSE component to support the basic requirements of the application;
- Performance of the Target System as a result of APSE-produced software:
  - -- should be time efficient for time-critical applications;
  - -- should be space efficient for space-critical applications;
  - -- should be time and space efficient for time- and space-critical applications;
- Performance of the host development system and/or the field service system;
  - -- development efficiencies;
  - -- failure/recovery (support/efficiency);
- Tool capacity:
  - -- APSE components must be able to support the hardware/software capacity constraints of an application;
- Distributed and/or multiprocessor systems;
- Application-specific benchmarks should be required.

2.3.1 Background

The evaluation of APSEs (software engineering tools) should be specific to the environment in which they will be used. The purpose of an APSE is to assist in the development of software systems. Therefore, the end user will be more apt to make use of the E&V technology if it provides a means of identifying the appropriateness of an APSE (tool-set) to that user's specific application. The E&V effort needs to assist in the identification of tool properties that will reflect application-oriented areas of interest -- i.e., efficiency issues, capability issues, performance issues, etc..

2.3.2 Motivation

(a) Capability

An APSE is a collection of software engineering tools. The purpose of building these tools is to employ them in the development of software systems for "mission-critical" embedded computer systems. Clearly, APSE contributes directly towards addressing the goals of these mission-critical applications. Therefore, it is important to evaluate the capabilities of an APSE specific to the application environment in which it will be employed.

(b) Performance (Target System)

Typical embedded-system applications are mission-critical and real-time. Often mission success is a function of time. In order to support the development of this application type, an APSE must adhere to the time and space constraints. Therefore, evaluation of an APSE to conform to the time-critical performance is extremely important.

(i) Time Efficiencies: Time-critical applications are constrained as to the total CPU time which they can use. Because of this, appropriate APSE tools must be evaluated in order to determine how well the resultant product can meet time efficiency requirements.

(ii) Space Efficiencies: Space-critical applications are constrained as to the total amount of memory or allotted system partitioning which they can use or in which they can reside. Because of this, appropriate APSE tools must be evaluated in order to determine how well the resultant product can meet space efficiency requirements.

(iii) Time and Space Efficiencies: Combinations of demanding reaction time and stringent space constraints are a real criteria in battlefield scenarios. Mission success indeed rests on these constraints. In order to support the software development for these systems, the APSE must comply with their combined requirements. Therefore, time and space efficiency evaluation of an APSE is critical.

(c) Performance (Host System-Development/Field)

(i) Development Efficiencies: In any development environment costs are schedule drivers. In order to keep the development cost under control and comply with the schedule constraints, development efficiencies play a key role, therefore the requirement to assess APSEs for development efficiencies is important. APSE evaluation that measures the efficiency for developing applications will contribute significantly to programmer productivity (one of the important goals of the STARS Program).

(ii) Fail-Recovery (Support/Efficiency): Failure in software systems in the field is emphatically more critical than in hardware. Hardware failure is a single unit failure -- unless it is a design flaw--, whereas software failure encountered in the field is present in every system that has this software. It is not possible to recall every software system from the field. Indeed, it is critical to resolve the failure in the least amount of time possible. APSE evaluation of the support and efficiency of fail-recovery is significant.

(d) Tool Capacity

An important consideration when developing embedded application software is the effect on APSE requirements that hardware-specific and software-specific items have on E&V. The evaluation criteria should include the effect that a machine architecture can have upon APSE tools -- for example, how different memory sizes affect the generated code. There also needs to be a measure of the capacity of a tool with respect to system size. For example, does a compiler successfully handle the number of identifiers needed within its symbol table? Will a linker be able to handle the needed segmentation or needed overlays of code? Requirements must include a way to assess an APSE's capacity to produce code for various system configurations.

(e) Distributed and/or Multiprocessor System (Support/Efficiency)

The continued higher processing throughput demands due to the increasing data densities are forcing the use of distributed/multiprocessors/parallel-processors in embedded computer systems. In order to accommodate these stringent

requirements of applications, evaluation of distributed/multiprocessors software is of paramount importance. This issue not only needs addressing in E&V but in CAIS as well, and corresponding requirements must be formulated.

(f) Application-specific Benchmarks

In order to assist in evaluation of APSE tools for a specific application, benchmarks must be provided to supply a means of directly evaluating how well a tool performs, as well as meets application constraints. Benchmarks are essential in order to provide measurements of execution speed and space utilization in program development. Various "mission-critical" applications have different emphases on performance requirements. APSEs used in the development of these special applications have direct bearing on their performance requirements. Therefore, the application-specific APSE evaluation has a requirement for application-specific benchmarks.

2.3.3 Example Application Areas

Electronic Warfare (EW) Applications

The data processing needs of this type of system are detecting, identifying, and reporting. On the surface it may appear simple, but in reality it is not simple at all. EW is termed a mission-critical system. As such, any software developed for this application must comply with the dictates of time-critical features and space constraints. Ultimately, the software developed for these systems must aim for mission success. The APSE tools that apply to this area must meet the efficiency requirements of these systems, as well as assist in the efficient development of software for these systems.

Avionics Applications

The development of a navigation system is a real-time application that can benefit greatly from a tool-set that assists in the well-disciplined development of software. The reliability aspect of navigation systems is a prime concern. With systems becoming more complex, yet having to be more compact, the areas of tool

support and efficiency are extremely important. The use of software-testing tools is also gaining importance because of the robust quality that their use can give to the final software product.

2.3.4 Issues

Applications involving embedded systems must be concerned with efficiency. APSE tools need to be evaluated with respect to the efficiency of the software produced by them. For example, a compiler's efficiency is based not only on the number of lines of source code compiled per minute, but also on how efficiently the produced code executes when put into a specific application environment. In addition, the capability of a compiler may include not only how well it handles I/O, but how much program capacity it supports.

Production efficiency is another issue. How helpful is the tool-set with respect to time needed for system development? Evaluation criteria should include the turnaround time of software problem resolution once it is in the field (failure/recovery) and the reliability of these software changes. How well does it support the production of quality software?

Run-time support is another issue of interest to producers of embedded systems. The actual performance of run-time components is important to include in E&V of APSEs. Precision requirements can be examined as well as time constraint requirements on the use of run-time support components. For example, a math library routine can be evaluated for execution time vs precision needs. With this type of information available via APSE E&V, the application area can determine the proper trade-off between the two attributes.

2.3.5 Conclusion

It is a point well taken that not every application can be accommodated in the E&V task. But the specific area of embedded computer systems should be dealt with. How to implement the requirement to include evaluation criteria that address specific applications can occur first through solicitation of benchmarks to be included later in E&V

2.4 Requirements Relating to Evaluation of Simulation Support and Embedded Software Development

(1) to evaluate simulation support capabilities, and
(2) to evaluate capabilities for transitioning applications from host to target.

2.4.1 Background

Based on the obvious importance of supporting the development of mission-critical embedded computer software, it is recommended that two categories of APSE tools, largely ignored in the NBS Tool Taxonomy, be included in APSE E&V technology development plans. These two categories are simulation support and host-to-target transitioning support. Tools supporting these capabilities are needed for early prototype development of the embedded code, for moving the code to a target environment, for testing and verification, and for maintenance/modification of previously delivered code.

2.4.2 Motivation

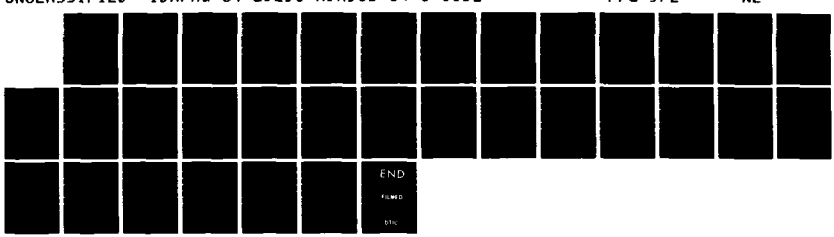
Potential interface issues are raised by the distributed simulation/development scenario alluded to above. Both synchronous and asynchronous communication and synchronization requirements exist, depending on the nature of the simulation and embedded system. Although CAIS 1.1 defers the question of distributed support, the question of compatibility with CAIS and the possibility of developing a distributed system or a multiprocessor simulation (e.g., a "Hardware-in-the-loop" simulation) arise. It is recommended that these issues be considered by the CAIS Working Group, and that corresponding requirements be formulated in the E&V Requirements Document.

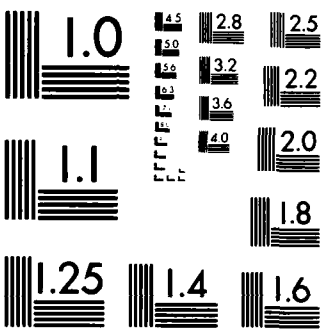
2.4.3 Examples of Desired Tools/Features

Examples of simulation support tools are

- discrete-event schedulers,

AD-A156 667 ADA (TRADEMARK) PROGRAMMING SUPPORT ENVIRONMENT (APSE) 2/2
EVALUATION AND VAL.. (U) INSTITUTE FOR DEFENSE ANALYSES
ALEXANDRIA VA V L CASTOR ET AL. DEC 84 IDA-M-34
UNCLASSIFIED IDA/HQ-84-29295 MDA903-84-C-0031 F/G 9/2 NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963-A

- numerical integration routines;
- tools to support user-interactions with graphical displays;
- tools to support synchronization between simulations (running on a host computer) and embedded software (under test on a target computer).

Examples of transitioning tools are:

- cross compilers;
- target machine loaders;
- target machine instrumenters;
- down loaders;
- bootable media creators;
- data file translators.

2.5 Requirements for the Evaluation of Inter-Tool Interfaces

(1) For each APSE tool evaluate the degree of coupling between it and other tools in the APSE.
(2) Identify the APSE tool-sets and the tool components that constitute each tool-set.

2.5.1 Background

APSE E&V results should assist the APSE developer in identifying tool-set(s) that may easily be integrated into an environment. For example, one may want to integrate a specific vendor's compiler into an environment because it meets certain code efficiency requirements. The evaluation of the compiler-to-APSE interfaces should provide information that will assist the APSE developer in determining the effort required to incorporate this compiler into his APSE.

2.5.2 Motivation

The degree of coupling between tools is determined by the following:

(1) Is data shared between tools (e.g., produced by one tool, used by another tool)?
(2) What the tool needs to know about the data.

(3) How the interface is realized.

(4) The number of tools with which a tool interfaces.

The use of the CAIS in itself is necessary but not sufficient to ensure tool transportability and interoperability. The following discussion presents four levels of inter-tool interface coupling and illustrates these levels by example.

(1) No coupling -- tools do not share data.

(2) Minimum coupling -- Example: editor and compiler. An editor that produces a source text file to be used by a compiler. This file is a standard CAIS text file conforming to standard text file formats.

(3) Medium coupling -- Example: compiler & configuration manager. The compiler accepts as input a simplified logical library designation which is used to select the necessary library objects. The compiler has no knowledge of the configuration management data used in a particular APSE.

(4) Tight coupling -- Example: integrated compiler and configuration manager. The compiler retrieves objects directly from an APSE's library using knowledge of the configuration management data to retrieve these objects.

The degree of coupling can be used to identify tool-sets (as well as other metrics, e.g., cyclic dependencies). One may want to discuss interfaces between tool-sets as well.

Other coupling-related issues:

- high degree of data sharing within an APSE may indicate a higher degree of APSE tool synergy;
- high degree of tool interface coupling may imply low tool transportability;
- interface implementation evaluation:
  (a) use of standard forms (e.g., DIANA, text files),

(b) Ada packages define interfaces -- will require tool recompilation when transporting tool,

(c) tool interface table-driven -- can be parameterized for different APSEs,

(d) highly integrated and requires extensive tool modification to transport.

3.0 GENERAL CONCERNS

Several of the Requirements Working Group discussions resulted in important industry input to the E&V Team, but these results were not formulated into specific *requirements during the workshop. This section details areas of industry input* that need to be considered by the E&V Team's Requirements Working Group. The areas include Near-Term Evaluation Requirements and Levels or Perspectives of APSE Evaluations.

Near-Term Requirement. *The working group agreed on the need for* evaluation techniques that could be applied early-on in the evolution of APSEs. This need addresses the fact that early APSEs and "Non-APSEs" may be more suitable for one project than another. Near-term evaluations would center on providing data to make these distinctions. Although not discussed at length at the workshop, one suggestion was to use the document "Criteria for the Evaluation of ROLM Corporation's Ada Work Center", prepared by V.L. Castor, as a baseline for near-term evaluations. This document presents several well-formulated questions that, when answered, would provide a database of useful evaluation information. Although the criteria are presented specific to the ROLM system, with a minimal amount of effort they could be abstracted and parameterized so as to be more generally applicable. Using the document in this manner would quickly provide an evaluation mechanism.

Another near-term need, which will become less important as APSEs develop, deals with the ability to interface other tool-sets. A developing APSE may not include all functionality at an early stage of development. The ability to interface a non-APSE tool to the APSE would then be important. A simple example of this need is an APSE without an editor. In this case the ability to interface the APSE compiler to existing editors is important.

Levels and Perspectives of Evaluation. The working group agreed that there exist two distinct levels/perspectives for APSE evaluation. At one level evaluations are performed on the individual components of the APSE, and at the other evaluations are performed on the APSE as a whole.

Component evaluations take the form of tool evaluations, functionality evaluations, and performance evaluations. Tool evaluations are those that examine tools independent of their functionality. All tools on an APSE need to be evaluated based on their interfaces with other tools, Ada packages, the CAIS, and users. The complexity of connection between tools and the number of tools with which another interfaces are example evaluations presented. Both examples illustrate the functional independence of tool evaluations.

Functionality and performance evaluations examine APSE components based upon their support of benchmark functionality or performance. Examples might include a benchmark of functionality that supports bibliographic references in documentation tools, or a benchmark performance for compiling a source program.

Evaluation of the APSE as a whole involves examining the synergistic effect of the APSE. These evaluations, rather than examining individual tools or functions, examine the way in which tools interact at a global level. Examples of this include evaluating the consistency of user interface command names, argument formats, options formats, and error display formats. The desire is to have consistency among the tools of an environment.

4.0 REQUIREMENTS DOCUMENT OUTLINE

4.1 GENERAL INTRODUCTION

Purpose of E&V Task
Purpose of Requirements Document
Whom this Document is For
What it is to be Used For
Scope
Goals
Near-Term
Long-Term
Assumptions
Definitions and Acronyms

4.2 APPROACH

Objective, Repeatable, and Reliable Tests
Data Base of Results That Can be Made Available to Community
Reference Manual for E&V Tests
Required Resources
Guide Books to Using E&V Tests
Intended Use of Results
Evaluators
Data Base Users
User Feedback to E&V Team
Risks and Cautions in Using E&V Results
Organization of E&V Tests
Monitor Field Use of APSEs
Policies and Procedures

4.3 PRODUCT QUALITY GUIDANCE

Overall Criteria/Requirements
Criteria for Tools and Methods
Test Programs
Self-contained
Independent
Can be Grouped to Run as a Module
Do not Require Factoring to Use
Self-Checking Where Possible
Documented as to Purpose, Inputs, Outputs, and Results
Are Applicable to Generic APSEs
Checklists
Questionnaires
Configuration Management
Baselines must be Established
Version DescriptionDocuments
Established Procedures for Reporting Problems
Status Accounting on all Baselined Units
Quality Control of E&V Products
Beta Testing

4.4 APSE E&V Requirements
APSE Functional Taxonomy
Functions
Attributes including: "ilities"
human factors
documentation
CAIS/KAPSE E&V
Macroscopic Issues
Consistency of APSE
Ease of Accomplishing Tasks
Composition of Tools to Perform a Task
"ilities"
Configuration Control of APSE by Sponsor
Monitor Field Use of APSE
Resources Needed to Use APSE
Implementation Dependencies

4.5 SUMMARY/CONCLUSIONS

APPENDIX D

APSE EVALUATION REFERENCE MANUAL WORKING GROUP REPORT

APSE EVALUATION AND VALIDATION (E&V) WORKSHOP AIRLIE, VIRGINIA

April 2-6, 1984

APPENDIX D

APSE EVALUATION REFERENCE MANUAL WORKING GROUP REPORT

TABLE OF CONTENTS

I. E&V REFERENCE MANUAL WORKING GROUP OVERVIEW D-2

1.0 List of Members D-2
2.0 Goal D-2
3.0 Accomplishments D-2

II. E&V REFERENCE MANUAL DOCUMENT D-3

1.0 EXECUTIVE OVERVIEW D-3
1.1 Rationale D-3
1.2 Document Overview D-5

2.0 SCOPE AND PURPOSE D-6
2.1 Synopsis of E&V Plan D-6
2.1.1 E&V Objectives D-6
2.1.2 Technology Development D-7
2.2 Relationship of Documents D-8
2.3 Users and Uses of the Reference Manual D-10
2.3.1 User Model D-10
2.3.2 Usage Cross Reference D-12

3.0 GLOSSARY D-13
3.1 Glossary of Acronyms D-13
3.2 Glossary of Terms D-13

4.0 REFERENCES D-14

5.0 HOW TO USE THE REFERENCE MANUAL D-15
5.1 Structure of the Evaluation Reference Manual D-15
5.2 Accessing Information D-15
5.2.1 Conducting an Evaluation D-16
5.2.2 Determining Evaluation Criteria D-16
5.2.3 Locating References to Evaluation Reports D-17

6.0 APSE EVALUATION D-18
6.1 Introductory Paragraphs D-18
6.2 Format for a Detail Page (Evaluation Item) D-18
6.3 Overall Format for a Detail Page D-19
6.4 Near-Term/Long-Term Considerations D-20
6.5 Transition for Hard Copy to a Mixed Document D-23
6.6 APSE Global Evaluation D-24

I. E&V REFERENCE MANUAL WORKING GROUP OVERVIEW

1.0 REFERENCE MANUAL WORKING GROUP MEMBERS

Chairperson -- John (Jack) Kramer, Institute for Defense Analyses

Members:

Paul Dobbs, General Dynamics
Bud Hammons, Texas Instruments/North Texas State
Rick Long, Air Force Wright Aeronautical Laboratories
Ronnie Martin, Georgia Institute of Technology
John Reddan, Syscon Corporation
Ray Sandborgh, Sperry Corporation
Carl Schaefer, Intermetrics, Inc.
Jim Winchester, Hughes Aircraft Company

2.0 GOAL

- Relationship to Requirements Document (and Guidebook)
- Contents (long and short)
- Purpose and audience.

3.0 ACCOMPLISHMENTS

- Reference Manual outline
- Draft contents for each introductory section
- Detail page format and example
- Solution to volume and E&V time
- List of issues identified.

II. E&V REFERENCE MANUAL DOCUMENT

Note: This Appendix is not intended to be a complete document; each Section needs significant additional detail. It is intended to be used by the ultimate APSE Evaluation Reference Manual authors to provide initial guidance and ideas for the final document; it is to be used by the authors of other E&V documents so they can better understand what the Reference Manual will contain.

1.0 EXECUTIVE OVERVIEW

This section presents the rationale for the development of E&V technology and an overview of the APSE Evaluation Reference Manual.

1.1 Rationale

Why E&V Technology ...

The evaluation and validation (E&V) technology is being developed to provide the capability to perform assessments of Ada Programming Support Environments (APSEs) and their components and to determine conformance to appropriate standards.

Why Ada ...

Ada was developed to be the standard high order programming language for use in DoD mission-critical computer systems. The rationale for the development of a single language was principally based upon the benefits to be derived from the ability to share personnel, technology, facilities, software components, and other resources between projects and Services. A secondary benefit of having a single language was anticipated from not having to design a new language, compiler, and associated tools for each major new weapons systems.

Why APSE ...

Early in the Ada development process, it was recognized that the benefits derived from a common language could be increased substantially by the

development of an integrated system of software development and maintenance tools in support of the life-cycle use of that language. Hence, the Ada Progamming Support Environment (APSE).

Why CAIS ...

The Common APSE Interface Set (CAIS) has been developed to ensure the interoperability of data and the transportability of tools between conforming APSEs. This interoperability and transportability (I&T) is of utmost importance if the benefits of a common language and its associated programming support environments are to be realized without adopting only one APSE. It must be relatively easy to move tools and project data bases between APSEs.

Why E&V ...

The evaluation of an APSE or selected components and the validation of components where standards exist provide information to decision-makers on the suitability and effectiveness of that APSE or the E&V'd components for the given application. Validation methods are applied to determine if the item being examined conforms to required standards. Evaluation techniques are applied to assess qualities for which no standard exists (or for which a standard exists but no validation capability is available). The information obtained by applying the E&V technology to a project's APSE is essential to the assessment of the risk involved in the program development or post-deployment support, as appropriate.

State of E&V Technology ...

The E&V technology described in this document and associated documents is an evolving technology. The ultimate goal is to develop the technology necessary to perform (1) formal validations of conformance to appropriate standards, and (2) objective evaluations of the other qualities for which only subjective evaluations are possible today.

Available Documentation ...

The E&V Plan gives a much more detailed introduction to the effort. Other relevant references can be found in Section 4 of this Appendix.

1.2 Document Overview

The APSE Evaluation Reference Manual is organized as follows:

Section 1: EXECUTIVE OVERVIEW: Section 1 presents the rationale for the development of the E&V technology and an overview of the APSE Evaluation Reference Manual.

Section 2: SCOPE AND PURPOSE: Section 2 is a more detailed introduction to the Reference Manual. It presents a synopsis of the E&V Plan, an overview of the relationship of the E&V documents to the Evaluation Reference Manual, and a description of the possible users/uses of the Evaluation Reference Manual.

Section 3: GLOSSARY: Section 3 presents a glossary of acronyms and definitions.

Section 4: REFERENCES: Section 4 presents a list of references that are used within this document and where to get them.

Section 5: HOW TO USE THE REFERENCE MANUAL: Section 5 presents information that explains to the reader how to use the APSE Evaluation Reference Manual effectively.

Section 6: APSE EVALUATION: Section 6 presents index pages that direct the reader to detail pages which list the evaluation techniques to be applied. It also provides references to the tests and test scenarios that are found in the E&V Guidebook.

2.0 SCOPE AND PURPOSE

This section provides a fairly detailed introduction to the APSE Evaluation Reference Manual. It presents a synopsis of the E&V Plan, an overview of the relationship of the E&V documents to the Reference Manual, and a description of the possible users and uses of the Reference Manual.

2.1 Synopsis of E&V Plan

The overall goal of the E&V Task is to develop and provide to the community the technology to be used in the evaluation of APSEs and the validation of components of APSEs where standards exist. In order to accomplish this goal, eleven specific objectives have been identified.

2.1.1 E&V Objectives

(1) Develop requirements to be used as guidance in the APSE E&V effort.
(2) Develop an APSE E&V classification schema.
(3) Identify and classify APSE components.
(4) Develop APSE evaluation capability where no formal standards exist.
(5) Develop APSE validation capability where there are formal standards. In addition to MIL-STDs, validation procedures will be developed for ANSI and ISO standards where appropriate.
(6) Monitor the Formal Qualification Testing (FQT) of DoD-owned APSEs.
(7) Develop evaluation and validation tools and aids. These include test sets, test scenarios, data reduction capability, and other means of automated support.
(8) Develop procedures for implementation of E&V technology based upon E&V requirements, APSE standards, evaluation criteria, validation capability, and existing E&V tools and aids.
(9) Provide initiative and a focal point with respect to APSE E&V.
(10) Solicit industry and academia participation in the E&V Task. One of the methods to be used is E&V workshops.
(11) Promote community use and acceptance of the E&V effort.

2.1.2 Technology Development

In order to properly understand the requirements, identify or develop appropriate standards, classify an existing APSE and then evaluate it, an appropriate taxonomy must be developed for the APSE and its components. As a starting point for a taxonomy, the NBS (Ref. H1) and Sperry (Ref. S1) taxonomies were used. These taxonomies were reviewed to determine their applicability to the E&V Requirements and other documents. Based on this analysis, the E&V Taxonomy Document will be developed and refined.

In order to better understand the various aspects of Evaluation and Validation, an initial E&V Classification Schema was developed. In this schema, the term "Evaluation" represents a method of assessing the quality of APSE components for which no specific standard (i.e., MIL-STD, ANSI, etc.) exists, or for which a standard may exist but there is no known capability to measure conformance to that standard. The term "Validation" represents a method of determining conformance to a standard which is applicable to an APSE (e.g., MIL-STD-1815A, CAIS).

The determination of what methodology (i.e., evaluation or validation) to use is then based on whether a standard exists and whether a means of checking conformance to that standard also exists. Since different levels of conformance checking exist, validation methodology is partitioned into non-formal and formal techniques. There is, in fact, a spectrum of evaluation and validation ranging from purely subjective evaluation to some form of a formal validation.

For a single component of an APSE (tool, functional component, or whatever) more than one method may apply at the same time. In addition, some characteristics can at this time only be measured subjectively, such as the quality of the error messages produced by a compiler. As an example, consider the generic tool "compiler" (or, if you prefer, the compilation function): Ada compilers must be validated for conformance to MIL-STD-1815A; but at the same time, a compiler can be measured objectively for such characteristics as compilation speed and speed of the produced code.

For such a component, the final evaluation would be a combination of all the measurements against the component, including both evaluations and validations. A comparison of two or more such components must be based on all the factors the person making the comparison deems important. Since these factors vary, this manual does not attempt to place weights on the various tests which may be levied against a component.

2.2 Relationship of Documents

This paragraph describes how the APSE Evaluation Reference Manual is related to the:

- Evaluation and Validation Plan
- Evaluation and Validation Requirements Document
- Evaluation and Validation Guidebook, and the
- Classification Schema or Taxonomy.

The Evaluation and Validation Plan identifies the requirement which led the AJPO to establish the E&V effort, provides a brief background and history, establishes the scope and primary purpose of the E&V effort, and sets the schedule for the initial and subsequent deliveries of the APSE Evaluation Reference Manual and other E&V documents.

The Classification Schema or Taxonomy is the structure used within the Reference Manual to organize the requirements.

The Evaluation and Validation Requirements Document defines the requirements which will guide the E&V Team in the development of E&V Technology and in particular those which the Evaluation and Validation Reference Manual meet, in addition to those specified in the Evaluation and Validation Plan.

The Evaluation and Validation Guidebook provides the detailed information on how to use the tools and techniques identified in the APSE Evaluation Reference Manual.

These relationships are illustrated in Fig. D-1.

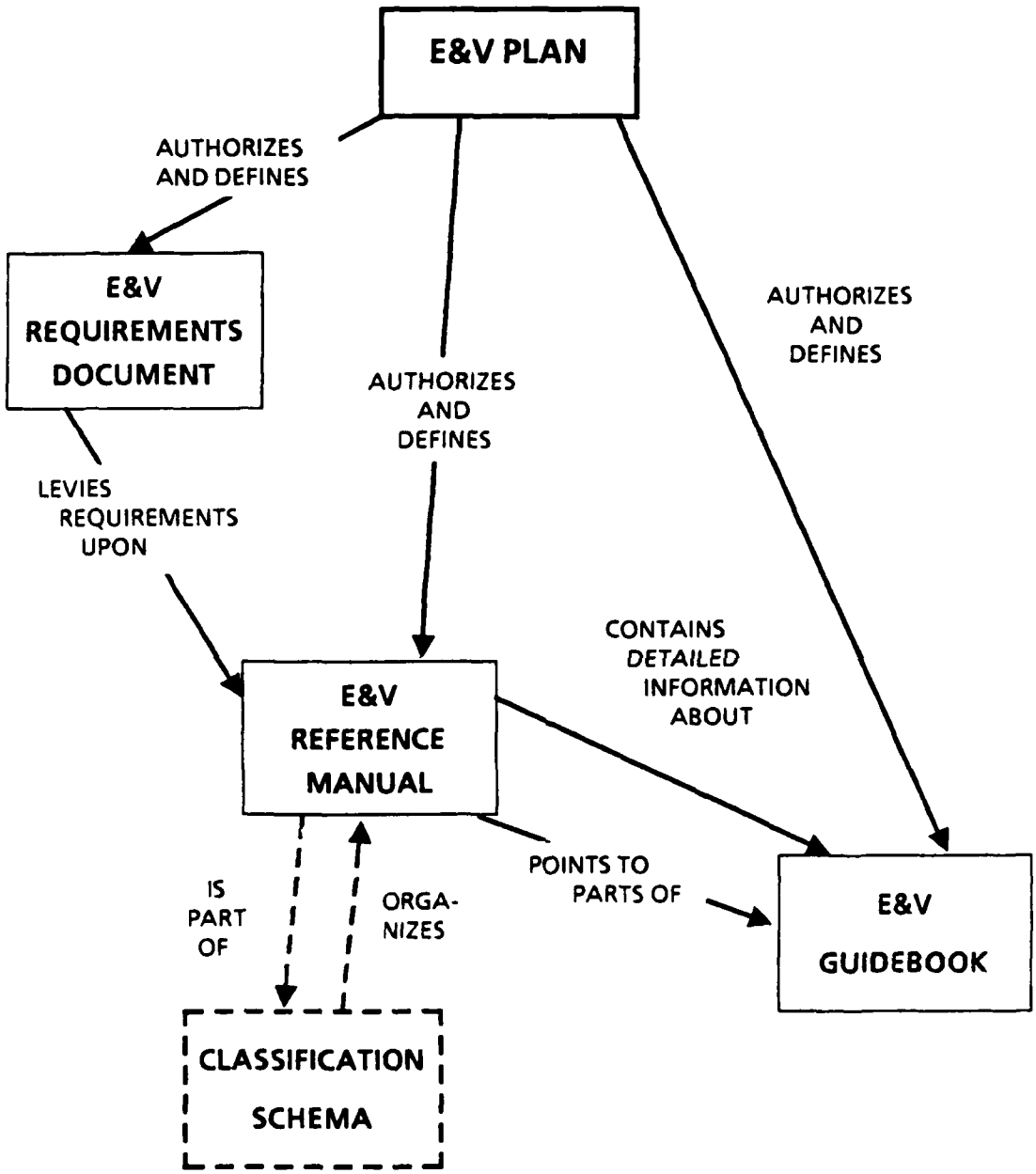


Figure D-1. Relationship of Documents

2.3 Users and Uses of the Reference Manual

The purpose of this section is to identify a model of the users of the E&V Reference Manual, the information that each user might extract from the Manual, and its application, and indicate in which chapters the information is located. Subsection 1 will list the user classifications of the model, and subsection 2 will describe the information extracted, give examples of its usage, and provide references to the applicable chapters of the Reference Manual.

2.3.1 User Model

Table D-1 shows the five user classes identified by the model of E&V Reference Manual users. This model was used to assist in identifying the issues and granularity of the topics discussed within the Manual. A review of this and the following subsections will allow the reader to determine the most useful sections for his or her consideration.

NAME	DESCRIPTION
TOOL PROCURER	WHO: DoD or IndustryProject Manager, e.g., Project Manager at XYZ Corp, Contract Manager at Avionics Lab. CHARACTERISTICS: High-level manager (in program or project); May not have a technical background Lacks the time to perform a comprehensive study
TOOL USER/SELECTOR	WHO: User of a potentially E&V'd tool or APSE, e.g., Software Development Staff Member CHARACTERISTICS: Technical user of tool or APSE on a specific machine(s)
TOOL DEVELOPER	WHO: Engineer evaluating feasibility (technical and economic) and/or developing tool definition and requirements, e.g., Chief Designer of Tool XX CHARACTERISTICS: Determines if tool is feasible and economically viable Sets tool requirements and development standards Has a software/technical background
E&V TECHNOLOGY USER	WHO: the applicator of E&V technology e.g., QA department of tool implementor the designated DoD validator (evaluator) potential tool buyer using as a selection criteria CHARACTERISTICS: Anyone applying E&V technology to the evaluation of APSEs or tools
INVESTOR	WHO: funder of developer or user of E&V technology e.g., Congress, AJPO, any funding agency High-level manager investing up-front money (PM or VP) Potential user of product developed with E&V'd tools and/or APSEs CHARACTERISTICS: A person interested in E&V technology in general, as opposed to its application to a specific tool(s) or APSE(s) A direct or indirect controller of monies

TABLE D-1. USER MODEL CLASSES

2.3.2 Usage Cross Reference

Table D-2 shows the information needed by each user class of the model, the Reference Manual chapters which contain it, and the uses to which the information will be put.

USER	INFORMATION EXTRACTED	USAGE	APPLI-CABLE SECTIONS
TOOL PROCURER	What benefits are derived from E&V? What are the high-level procedures for using E&V technology?	Uses the technology to assist in make/buy decision. Set a course of action for procurement, including assignment of personnel (Tool Sector, EV Technical User) to apply or study tools or APSEs.	2-5
TOOL USER/ SELECTOR	What are the global issues and functional components of a tool or APSE?	Use to locate reports on E&V'd tools and APSEs. Find metrics for tool assessment. Determine if E&V technology must be applied (then see E&V Technology User).	2-6
TOOL DEVELOPER	What are the global issues and functional components of a tool?	Use to develop list of evaluation items applicable to tool to be developed (will be formed into requirements or guidelines).	2-6
E&V TECH-NOLOGY USER	What are the global issues and functional components of a tool or APSE?	Use to index the Guidebook for details on how to perform the E&V tests.	2-6
INVESTOR	What is E&V? What are the goals, benefits, and background of E&V?	Use to determine the benefits of E&V and to decide whether to allocate funds.	1-2

TABLE D-2. USAGE CROSS REFERENCE

3.0 GLOSSARY

[NOTE: This is an example to show format and content. The final Glossary will be completed when the Reference Manual is completed.]

The Glossary has been divided into two parts - a glossary of acronyms and a glossary of terms. Each one includes the page on which it is explained or first used in a defining context.

3.1 Glossary of Acronyms

[Acronym -- Expanded
Page of good example of use]

APSE -- Ada Programming Support Environment
Page D-3

E&V -- Evaluation and Validation
Page D-4

3.2 Glossary of Terms

[Term -- Definition
Page of good example of use]

Ada -- a high-level computing language
Page D-3

Evaluation -- a method of assessing the quality of APSE components for which no specific standard (e.g., MIL-STD, ANSI) exists; or for which a standard may exist, but there is no known capability to measure conformance to that standard.
page D-7

4.0 REFERENCES

[NOTE: When the final Reference Manual is produced, this should be an exact extract from the E&V Team Project Reference List.]

[Ref.#] Author(s), Title, Publication, Page, Agency, Date, Vol.
How to get document and cost of document.

[B1] Buxton, J.N., "Requirements for the Ada Programming Support Environment (APSE)", U.S. Dept. of Commerce, Feb. 1980, Vol. xx. [How to get and cost to be filled in]

[C1] Castor, Virginia L., "Evaluation and Validation (E&V) Plan, Version 1.0," AFWAL/AAAF, Wright-Patterson AFB, OH 45433, 30 November 1983. [How to get and cost to be filled in]

[H1] Houghton, Raymond C., "A Taxonomy of Tool Features for the Ada Programming Environment (APSE), U.S. Department of Commerce, National Bureau of Standards, Institute of Computer Sciences and Technology, Center for Programming Science and Technology, Washington, D.C. 20234, Dec. 1982, NBSIR 82-2625 [How to get and cost to be filled in.]

[R1] Roberts, Teresa Lynn, "Evaluation of Computer Text Editers," Ph.D. Dissertation, Department of Computer Science, Stanford University, Stanford, CA, 1980: Available from University Microfilms, Ann Arbor, Michigan. [Cost to be filled in.]

[S1] Sperry Corporation, "A Tool Taxonomy," [Complete reference and how to get and cost to be filled in.]

5.0 HOW TO USE THE REFERENCE MANUAL

This section will assist you in using the APSE Evaluation Reference Manual.

Your purpose in using this manual may be:

(1) To evaluate an APSE or some component;

(2) To determine the criteria against which an APSE or a component will be evaluated (for example, in the course of designing a component);

(3) To locate references to reports of evaluations of APSEs or APSE components.

This manual is structured in such a way that, regardless of your purpose, you will access the information you need in the same way.

5.1 Structure of the Evaluation Reference Manual

The structure of the Evaluation Reference Manual allows users to easily focus their search for the APSE evaluation capabilities of interest.

The structure consists of a network of index entries that hierarchically relate APSE functional components into expanding levels of detail. Each index entry may contain references to other index entries or to detail pages containing information on evaluation items. Each detail page describes the rationale for an evaluation item and where the appropriate tests can be located in the Evaluation Guidebook. The starting point for the classification structure is the complete APSE.

Detail pages exist for both single component or tool evaluation items as well as for global evaluation items dealing with the interface between two or more tools.

5.2 Accessing Information

To access information, you must have a starting point. If you are interested in an entire APSE, then the component "APSE" is your starting point. If you are interested in an editor, then the component EDITOR is your starting point.

Find the index page corresponding to your starting point. This index page will refer you to detail pages or to other index pages or to both. You follow recursively whatever page references are relevant. The actual information you need will be in the collection of all detail pages to which you have been referred.

5.2.1 Conducting an Evaluation

If you are conducting an evaluation, you will be interested in the evaluation method given for each evaluation item in the detail pages. The detail page will also refer you to the Guidebook for procedural information on conducting tests as well as for the availability of tests of the appropriate type.

If you are conducting an evaluation of an entire APSE, then you should follow the reference (from the index page for the component APSE) to the index page for inter-tool interfaces, as well as the references for the individual APSE components.

In conducting an evaluation, you are dependent on the documentation of the tool developer or APSE developer for information about which subcomponents the product has. If the documentation indicates that a subcomponent is present, then you follow the corresponding index-page reference or detail-page reference. If the subcomponent is absent, then this absence is noted in the evaluation report.

5.2.2 Determining Evaluation Criteria

If you are trying to determine the criteria against which a component or an entire APSE will be evaluated, you are interested in the same information as you would be if you were conducting an evaluation, except that the information contained in the Guidebook would probably be below the level of detail that you need. Each subcomponent you encounter in traversing the index pages to the detail pages can be interpreted as a desirable functional capability of an APSE or APSE component.

5.2.3 Locating References to Evaluation Reports

If your purpose is to locate references to evaluation reports, you again access information by traversing the index pages to the detail pages. But you do not need to pursue the references (in the detail pages) to the Guidebook. Instead, each detail page will refer you to appropriate entries in the References section of this manual, which will tell you where to find reports on evaluations of the component you are interested in.

6.0 APSE EVALUATION

This section provides the detailed information the user needs to conduct an APSE evaluation or component/function evaluation or validation. It also explains the rationale and description of what must be done along with where to find the tests, test scenarios, and evaluation forms.

6.1 Introductory Paragraphs

The "detail pages" are intended to provide a detailed description of the evaluations (and validation capabilities) available at any given time for an APSE Functional Component ("cell" in the taxonomy) or a "canonical tool". The detail pages will provide overview information for each applicable evaluation that includes:

(1) The name of the evaluation;
(2) Purpose of the evaluation;
(3) Information reported;
(4) Very brief description of the nature of the evaluation;
(5) Who administers evaluation;
(6) Reference to applicable standards, supporting documents, the Guidebook, etc.;
(7) Resource requirements for assessment.

6.2 Format for a Detail Page (Evaluation Item)

(1) The name of the evaluation
   -- a descriptive label indicating the attribute of an APSE Functional Component being assessed.
(2) Purpose of the evaluation
   -- a very brief explanation of why the reader might care about the evaluation in question (this may, of course, seem redundant in the presence of a highly descriptive name for the evaluation item).
(3) Information reported
   -- the kind of information provided by the evaluation:

- category A-E
- pass/fail (in the case of a validation)
- presence/absence
- assessments by subjective scaling
- quantitative data (e.g., graph of response time to # users)
- formal verification
- description of mechanism/capability via empirical experience of evaluation or exploitation of documentation.

(4) Very brief description of the nature of the test

- automated test sequence
- monitored experiment
- synopsis of documentation
- questionnaire [respondent(s), statistics(?)]
- metrics.

(5) Who administers the evaluation

-- organization assigned
-- contractor/manufacturer
-- reader can perform with appropriate documentation.

(6) Reference to applicable standards, the Guidebook, and (possibly) other sections of the Reference Manual.

6.3 Overall Format for a Detail Page

APSE Functional Component Name

- Brief Description
- Evaluation Item 1
- Evaluation Item 2
  - .
  - .
  - .
- Evaluation Item n
- Prior Experience References.

Note: Some evaluation items may be considered sufficiently close in relation that they are organized under a descriptive heading for presentation purposes.

6.4 Near-Term/Long-Term Considerations

In order to provide a transition from the beginning of the E&V technology effort to a mature evaluation technology, given our current understanding of tools by commonly understood canonical classes such as editors or compilers, we have adopted a "data structure" approach to the organization of the Reference Manual. Further, the current absence of a provisional taxonomy to guide classification dictates that a fixed organization of the Manual is premature.

With a sufficiently adaptable structure, the Reference Manual can conform to the taxonomy as it evolves and also adapt itself to new canonical tools as they are conceived. The list of canonical tools at any given time is an important entry point for users of the Reference Manual, as users will likely continue to categorize a useful ensemble of features (APSE Functional Components) as an identifiable tool.

One way to view the relation of the collection of canonical tools and the suite of AFC (APSE Functional Components) that exists at a given time is as many layers of abstraction. Some canonical tools may be expressed for evaluation purposes as unions, intersections, and interfaces among (less abstract) canonical tools. Our data structure approach is geared to support the expression of such dependencies. In any case, the "bottom layer" in this structure is the assessment descriptions for the set of AFCs. There is effectively no upper boundary to this abstraction save that of the general notion "APSE", which is intentionally left ambiguous.

The near-term picture of the Reference Manual structure is provided in Figs. D-3 and D-4. A canonical tool is described, for assessment purposes, by a provisional set of direct references to tangible assessment procedures in the Guidebook, as well as to documentation.

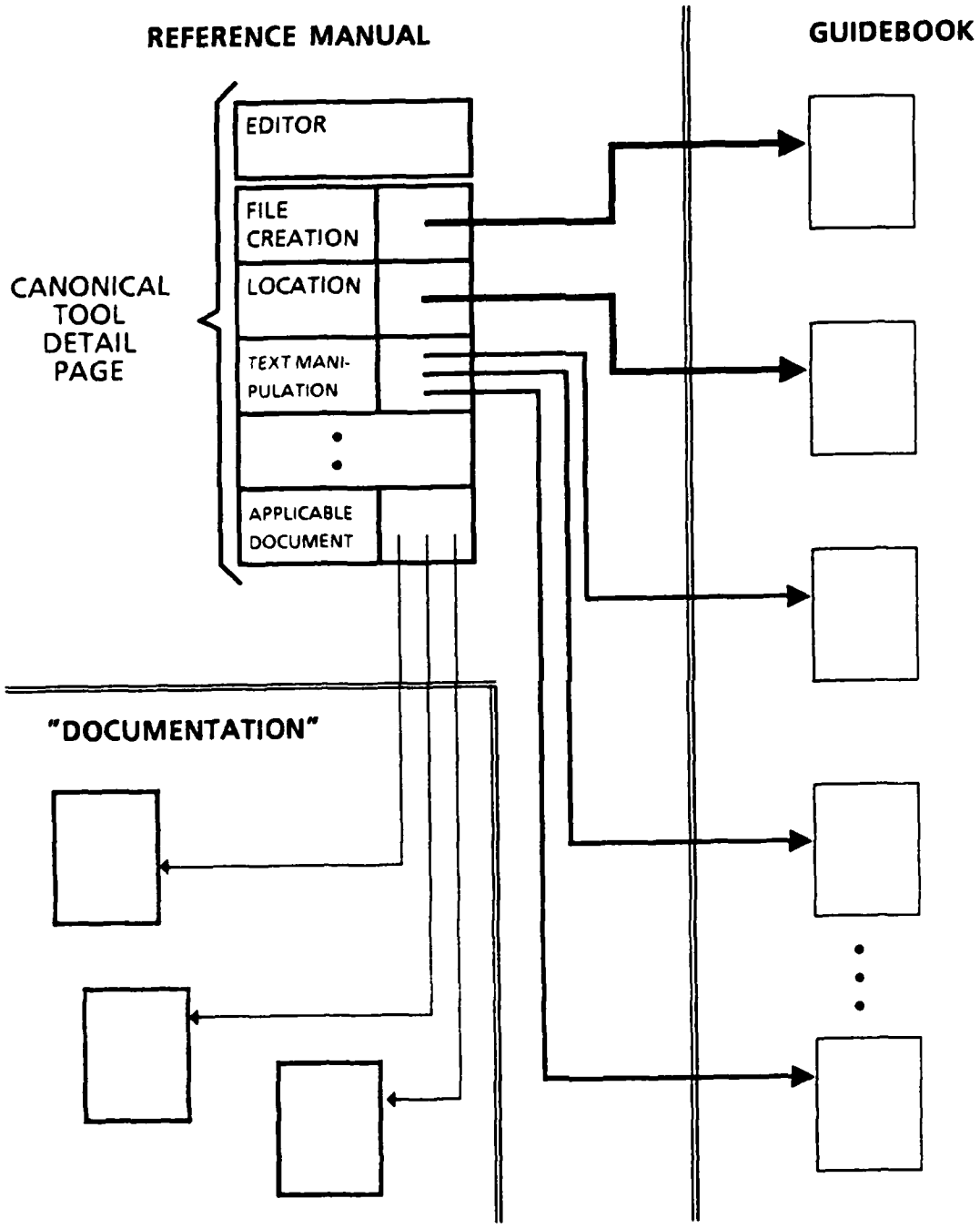
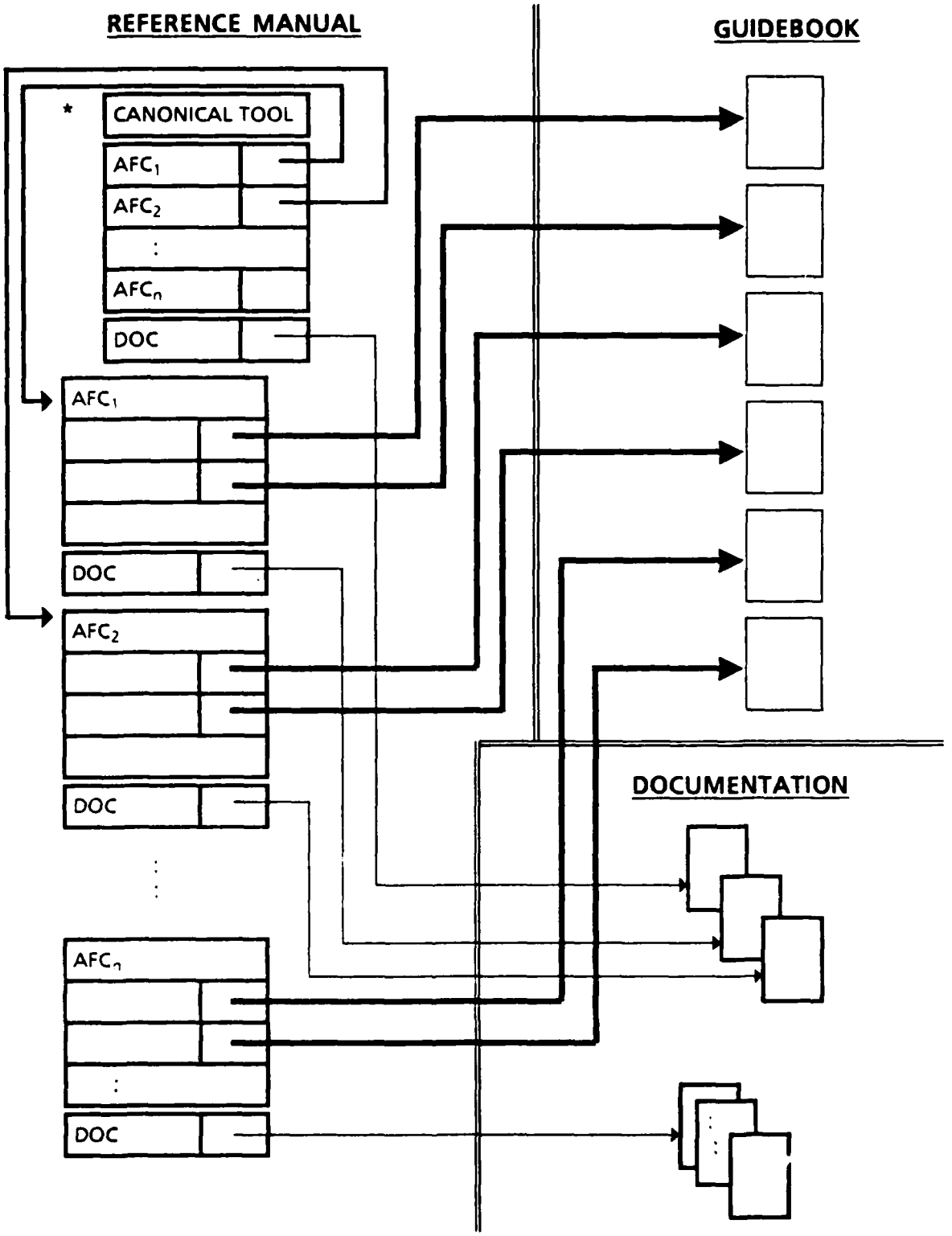


FIGURE D-3. Near-term Considerations



*Note: This relationship is also established by the classification of a particular tool with respect to the taxonomy.

FIGURE D-4. Long-term Considerations

As the taxonomy is elaborated we anticipate the structure of the taxonomy to be implicitly represented in the Reference Manual's structure. The AFC detail pages will refer to the Guidebook, while more abstract canonical tools will refer to AFCs to express assessment needs.

The question of how to assess and classify new tools will occur early in the evolution of the Reference Manual. A first approach is to use a questionnaire (or questionnaires) to gather the needed data.

- How to begin with "canonical tools" and transition to a functional taxonomy of APSE capabilities;
- Short term -- probably a list of commonly occurring "clusters" of capabilities (e.g., an editor, compiler);
- Long term -- non-trivial tools considered explicitly as a suite of capabilities found in the taxonomy.

6.5 Transition from Hard Copy to a Mixed Document

At the beginning of this effort the Reference Manual should be manageable as a paper document. At some threshold the Reference Manual data structure should be made machine-readable. The portion replicated in "hard copy" should encompass rather high levels of abstraction that may express global concerns about an APSE as well as the current suite of canonical tools. Thus an argument can be made that computer implementation of the Reference Manual data structure should occur early, and that the paper version be limited as practicality dictates.

It is important to manage the complexity of using the Reference Manual system. One potential development is that the detail pages may ultimately have a one-to-one mapping into the Guidebook assessment procedures.

6.6 APSE Global Evaluation

This manual is oriented towards evaluating specific functional components of an APSE that are supported by individual tools. Users will also desire to evaluate the synergism between a collection of tools. The APSE Global Evaluation Categories and associated items might include:

- **Command Language Interface**
  (abbreviating, method of passing parameters, icon placement/selection,...)
- **Data Storage Interface**
  (data storage mechanisms -- dbms, flat file,...)
- **Analysis Notation Interface**
  (type of graphics, text , notation model)
- **Analysis Techniques Interface**
  (analysis model, algorithms)
- **Data Transfer Interface**
  (appropriateness of tool, output data for tools, input data)
- **Procedural Interface**
  (can tool, steps mesh with tool, steps).

END

FILMED

8-85

DTIC