2003

SCIENCE AND TECHNOLOGY PEER REVIEW: ADVANCED TECHNOLOGY DEVELOPMENT PROGRAM REVIEW

By

Dr. Ronald N. Kostoff
Office Of Naval Research
800 N. Quincy St.
Arlington, VA 22217
Phone (703-696-4198)
Fax (703-696-4274)
email: kostofr@onr.navy.mil

Mr. Robert Miller
DDL OMNI Engineering, LLC
McLean, VA

Mr. Rene Tshiteya
DDL OMNI Engineering, LLC
McLean, VA

(The views expressed in this report are solely those of the authors and do not reflect the views of the Department of the Navy, any of its components, or DDL OMNI Engineering, LLC.)

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (FROM - TO)

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER 5b. GRANT NUMBER 5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER 5e. TASK NUMBER 5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME AND ADDRESS ,	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME AND ADDRESS ,	10. SPONSOR/MONITOR'S ACRONYM(S) 11. SPONSOR/MONITOR'S REPORT NUMBER(S)
12. DISTRIBUTION/AVAILABILITY STATEMENT ,	
13. SUPPLEMENTARY NOTES	
14. ABSTRACT	
15. SUBJECT TERMS	

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19. NAME OF RESPONSIBLE PERSON
a. REPORT Unclassified	b. ABSTRACT Unclassified	c. THIS PAGE Unclassified			19b. TELEPHONE NUMBER International Area Code Area Code Telephone Number DSN

TABLE OF CONTENTS

ABSTRACT

The science and technology (S&T) programs sponsored by the United States Department of the Navy (DoN) are divided into three major budget categories:

1) Basic Research (6.1)
2) Applied Research (6.2)
3) Advanced Technology Development (6.3)

In 1999, DoN commissioned an internal review of the 6.3 program. A thirty-one member review panel met for one week to rate and comment on six evaluation criteria (Military Goal, Military Impact, Technical Approach/ Payoff, Program Executability, Transitionability (to more advanced development/ engineering budget categories or acquisition), Overall Item Evaluation) for each of the fifty-five presentation topics into which the mid-$500 million per year 6.3 program was categorized. This report describes the review process, documents insights gained from the review, summarizes key principles for a high-quality S&T evaluation process, and presents a network-centric protocol for future large-scale S&T reviews.

KEYWORDS: peer review; technology development; evaluation criteria; network-centric; group-ware; future naval capabilities; program review; research assessment; technology assessment; research evaluation; military relevance; military payoff; military impact; technical approach; metrics, Science Court.

1. EXECUTIVE SUMMARY

The science and technology (S&T) programs sponsored by the United States Department of the Navy (DoN) are divided into three major budget categories:

1) Basic Research (6.1)
2) Applied Research (6.2)
3) Advanced Technology Development (6.3)

In 1999, DoN commissioned an internal review of the 6.3 program. A thirty-one member review panel met for one week to rate and comment on six evaluation criteria (Military Goal, Military Impact, Technical Approach/ Payoff, Program Executability, Transitionability (to more advanced development/ engineering budget categories or acquisition), Overall Item Evaluation) for each of the fifty-five presentation topics into which the mid-$500 million per year 6.3 program was categorized. This report describes the review process, documents insights gained from the review, summarizes key principles for a high-quality S&T evaluation process, and presents a network-centric protocol for future large-scale S&T reviews.

Overall 6.3 Program Results

For the evaluation criteria Military Impact, Technical Approach, Program Execution, Transitionability, and Overall Item Evaluation, distribution functions of numbers of programs vs. rating bands (Low, Medium, High) were presented. No systemic overall 6.3 problems were uncovered.

Programs Related to Future Naval Capabilities (FNC)

In 1999, the naval services had identified twelve FNCs that were deemed as high priority targets for development. For the evaluation criterion Military Goal, the number of programs related to each FNC with strengths of relationships above parametrically-varied thresholds was obtained. In addition, the number of programs related to multiple FNCs was calculated. All 6.3 programs were related to at least one FNC with a strength of relationship of Medium or higher, and 95% of the 6.3 programs were related to at least one FNC with a strength of relationship of High. Some 6.3 programs were related to as many as eight FNCs with a strength of relationship of Medium or higher, and a few 6.3 programs were

related to as many as four FNCs with a strength of relationship of High. Having this understanding of inter-relationships will be invaluable in helping the Execution Managers coordinate the program management and output among the IPTs.

Individual Program Results

The panel-averaged ratings for each 6.3 item for the six criteria were generated. These data were used to determine the aggregate relationships noted above. A regression analysis of the five component criteria against the Overall Item Evaluation criterion was performed, to determine which criteria had the most influence on bottom-line score (Overall Item Evaluation). Two criteria, Military Impact and Technical Approach, provided the bulk of the influence on the determination of bottom-line score. A model consisting of these two criteria predicted the bottom-line score to within two per cent. This is consistent with other large-scale reviews (DOE, 1982; Kostoff, 1997d).

Recommendations for Action

Numerical results were used to place the fifty-five 6.3 items in broad quality categories. Specific actions recommended for each item depended heavily on the comments from the reviewers, with special attention paid to the comments from the user/ customer representatives. In general, no corrective action was recommended for items that had good performance and execution, good transition potential, and strong relation to at least one FNC. Various levels of correction, including termination, were recommended for items that had the following characteristics:

- Insufficient commitment to transition
- "Core-Program" structure
  -Insufficient FNC focus
  -Insufficient demonstration focus
- Potential for high cost over-run

Insights gained from both the planning and conduct of the review should be of considerable value when conducting future large-scale 6.3-type reviews, and include the following:

1) Provision of detailed programmatic descriptive material to the panelists and audience before the review is very useful; its value could be enhanced by e-mail interchange between the presenter or facilitator and the panelists before the presentations to clarify outstanding issues and allow for more effective use of actual meeting time.

2) Appropriate use of Group-Ware could allow:

- Streamlining the review process with real-time data analysis and aggregation
- Remote reviewer participation, thereby minimizing travel and logistics problems
- More reviewers to participate in the process, producing a more representative sample of the technical community
- Reviewers to be selected for expertise in specific evaluation criteria only, thereby enhancing the credibility of each rating
- Sufficient expertise on the panel such that the Jury function (fully independent decision-making) can be separated from the Expert Witness function (potentially conflicted technical judgment and testimony)

3) When assessing and comparing quality of programs representing multiple disciplines, it is necessary to normalize. Evaluating all programs in one setting is an excellent way to accomplish this objective. Because of the realistic time constraints associated with a single-setting review, depth must be traded off for breadth. This trade-off is acceptable, as long as depth is evaluated by some means during the S&T operational management cycle.

2. OBJECTIVES AND GOALS OF REVIEW

2.1. Background

The science and technology (S&T) programs sponsored by the United States Department of the Navy (DoN) are divided into three major budget categories:

1) Basic Research (6.1)
2) Applied Research (6.2)
3) Advanced Technology Development (6.3)

These categories are reviewed periodically to insure that a high level of technical quality is maintained, and that their constituent programs are relevant and responsive to intermediate and long-term naval services' goals. Typically, the programs within these categories are reviewed either individually or in aggregate about some central technical or mission theme.

2.2. Major Review Objectives

In 1999, DoN commissioned an internal review of the total 6.3 budget category. The objectives of the review were twofold: technical quality control and military relevance quality control for the total budget category.

2.2.1. Technical Quality Control

For the total 6.3 program review, assessing technical quality meant addressing issues such as technical approach and potential payoff relative to alternate technologies, demonstrating achievement of technical targets on schedule and cost, and ability to transition to more advanced development/ engineering budget categories (or acquisition) if demonstration succeeds.

2.2.2. Military Relevance Quality Control

In 1999, the naval services had identified twelve Future Naval Capabilities (FNC) that were deemed as high priority targets for development. It was desired specifically to ascertain the relation between the existing 6.3 program and the FNCs, in order to determine the level of management attention required to insure that the program would evolve seamlessly toward better alignment with the FNCs.

2.3. Review Sub-Objectives

Supporting these two major objectives were four important sub-objectives that drove the timing and structure of the review:

- Identifying systemic problems;
- Identifying FNCs requiring additional management attention;
- Increasing awareness of all DoN S&T stakeholders of technology development criteria important to DoN S&T management; and
- Optimizing the S&T portfolio for total FNC satisfaction.

2.3.1. Identifying Systemic Problems

One sub-objective was to ascertain whether there were any systemic strengths or weaknesses that transcended individual program characteristics, and required higher-level management attention than would be necessary for individual program problems. Attainment of this sub-objective required that the individual programs be evaluated on as common and standardized a basis as possible. This normalization procedure necessitated that the total 6.3 budget category be evaluated in one setting, using common evaluation criteria, with the same panel.

2.3.2. Identifying FNCs Requiring Additional Management Attention

A second sub-objective derived from the management structure instituted to insure S&T program responsiveness to the twelve FNCs. An Integrated Product Team (IPT) was established for each of the twelve FNCs. Each IPT had broad representation from the S&T, requirements, and acquisition communities. Each IPT had the charter of developing S&T programs that would respond to its particular FNC. This second review sub-objective was to ascertain the magnitude and quality of the existing 6.3 program relative to each of the IPTs S&T responsibility areas, as a starting point for relating the total existing 6.3 program to the totality of programs required, and therefore to what new programs had to be established by each IPT. Simply put, this sub-objective was to determine the supply-demand imbalance (if any) of the present 6.3 program for each of the FNCs.

2.3.3 Increasing Awareness of All DoN S&T Stakeholders of Technology Development Criteria Important to DoN S&T Management

A third sub-objective related to the composition of the IPTs, since the membership was drawn from very diverse communities. It was desired to increase the IPTs' awareness of the S&T criteria that are important to DoN S&T management in the development of technology. Toward that end, the IPT Chairpersons were invited to participate directly in the review, and the other IPT members were invited to attend the review as audience.

2.3.4. Optimizing S&T Portfolio for Total FNC Satisfaction

A fourth sub-objective was to insure that technology portfolio development for the total 6.3 program was aimed at optimizing total FNC satisfaction. Achievement of this sub-objective required that the goals of each IPT be presented in one setting in a standardized manner, and the multiple application characteristics of each program be understood and appreciated. These complex interactions between technologies and capabilities also required a single setting for enhanced understanding.

3. STRUCTURE AND CONDUCT OF 6.3 REVIEW

3.1. Ground-rules of Review

A number of ground-rules were established for the 6.3 review at the outset. These rules are summarized below in Table 1.

Table 1. Summary of 6.3 Program Review Ground Rules.

No.	Ground Rule
1	All programs within the 6.3 budget category that received funding in Fiscal Year 2000 (FY00) would be included in the review
2	The taxonomy used for structuring the review presentations would be the most recent one also used for program selection and management
3	For logistics purposes, the review presentations would be limited to one week duration
4	Information Technology Group-Ware would be used where feasible
5	The principles of high quality program review would be followed wherever feasible. These principles have been summarized in Kostoff (1997b), and updated in Appendix 1.

The main elements of the 6.3 review were:

- presentations of the 6.3 program by the DoN S&T Execution Managers to an evaluation panel,
- ratings and comments by the panelists,
- analysis, interpretation, and recommendations by the review's operational managers, and
- final decisions by DoN S&T senior management.

Within this scenario, the three major foundational blocks were selection of the evaluation criteria, selection of the evaluation panel, and selection of a taxonomy for categorizing presentations.

3.2. Selection of Evaluation Criteria

The prime objectives, as stated above, were to evaluate technical quality and

military relevance of the 6.3 budget category, especially relevance to the FNCs. In addition, since the 6.3 budget category has an underlying demonstration and product motivation, it was desired to see how well the individual programs met these hard deliverable targets. Five component criteria were defined to address both the potential technical and military payoffs, and the probability that this potential would be realized. These criteria are:

- Military Goal (relevance of program to military target),
- Military Impact (probability of producing military product),
- Technical Approach (potential technical payoff using specific approach),
- Program Executability (probability that technical targets can be demonstrated on time and budget), and
- Transitionability (likelihood that development would go to higher budget category or to acquisition after successful demonstration).

These were the component evaluation criteria selected. The specific definitions used, and sample evaluation forms, are shown in Appendix 2 (the generic term 'item' used in Appendix 2 refers to the funded technology development represented by each of the fifty-five presentations). In addition to the five component criteria, a sixth 'bottom-line' evaluation criterion (Overall Item Evaluation) was used, as shown on the sample form. The purpose of this overall criterion was to account for any factors that the reviewers thought might be important in evaluating a particular program, but that were not included in the component criteria. As will be shown later, the five component criteria captured all the major factors that were used by the reviewers in arriving at their 'bottom-line' scores.

3.3. Selection of the Evaluation Panel

Evaluation panels for S&T programs are usually of two limiting forms. One type consists of personnel completely external to the program(s) being evaluated, and if such personnel are also experts in the program's technical area, this review is termed a peer review (NRC, 1998; USNRC, 1988). Typically (not always), when peer reviews are used, they tend to focus primarily on detailed technical issues, and secondarily on mission-relevance and management-related issues. The second type consists of personnel associated with the organization that manages the program(s); this review is termed an internal review. Typically (not always), when internal reviews are used, they tend to concentrate primarily on higher level

mission-relevance management-oriented issues, and secondarily on detailed technical issues.

It was decided to perform an internal review using naval personnel entirely with some ONR management representation, for the following reason. The second sub-objective described above (Identify FNCs Requiring Additional Management Attention) reflected a transition of the 6.3 program from having a major 'core-like' structure to being much more strongly aligned and focused toward the critical FNCs. This new structure enhances the role of the technology customer/ user in the S&T decision-making process. The panel composition, with its relatively high representation from the requirements community, reflected this shift in emphasis. Also, as will be discussed later, recommendations resulting from the review were strongly influenced by the views of the user community representation on the panel.

In addition, because depth was traded for breadth in the 6.3 review, it was believed to be more important to have personnel represented on the panel that had a breadth focus rather than a depth focus. The panel members were also required to represent a diverse group of naval organizations, since the evaluation criteria spanned areas of authority of different naval organizations.

Four types of reviewers were included in the panel. These were:

- The Executive Steering Committee, the senior managers of the Office of Naval Research (ONR)
- Representatives from the Marine Corps
- Representatives from the DoN S&T resource sponsor (OPNAV 911)
- Advisors
  - Representatives from the Operational Navy organizations responsible for setting requirements.
  - Department Heads from ONR

A total of thirty-one reviewers were on the evaluation panel. Their civilian and military ranks were high-level, mainly civilians drawn from the Senior Executive Service and active military drawn from the Flag (Admiral) level.

3.4. Selection of a Presentation Taxonomy

The FY00 6.3 program was estimated (from the vantage point of FY99) to eventually be between $500 and $600 million. To complete the presentations within one week (a necessary ground-rule due to logistics considerations), about ten presentations per day seemed to be a reasonable limit. There were a couple of options for dividing the 6.3 budget category into separate presentations that would allow sufficient material to be shown for credible criteria evaluation. For the review, it was decided to use the taxonomy by which recent programs were selected and managed. This resulted in fifty-five separate presentations.

3.5. Conduct of the Review

With these foundational review blocks in place, the review proceeded as follows. A letter from the Chief of Naval Research was sent to all the major participants (presenters, reviewers, audience) initiating the review process. The letter included guidelines to the presenters (6.3 program Execution Managers) for generating canonical vugraphs that would address each of the evaluation criteria. The presenters generated the vugraphs (and backup material), and posted password-protected copies on the Internet a few weeks before the review. This allowed the reviewers and audience to become familiar with the fifty-five 6.3 programs before the actual presentations.

In parallel with the dissemination of background material, and logistics to prepare for the actual presentations, a Group-Ware software package was developed to help streamline the review process. This package would document the information flow from data entry of the reviewers' ratings and comments to final display of the results at the Executive Session at the end of the review. Time constraints did not allow a fully tested Group-Ware package to be implemented at the review, and only a portion of the capability was actually utilized. The package that was completed eventually, and processes in which it could be imbedded, offer the capability of a much enhanced peer or internal review approach. The software package is described in Appendix 3. A network-centric review process that would utilize this package, the experience of the 6.3 review and previous reviews, as well as reasonable extrapolations from these experiences, is described in Appendix 4.

The presentation sessions were classified at the SECRET level, and therefore no technical details will be presented in this report. The first segment of the presentation sessions consisted of the Chairpersons of the IPTs describing the scope and objectives of their FNCs. Because of the synergistic and symbiotic

nature of many of the FNCs (e.g., Information Distribution contributes to Missile Defense, Autonomous Operations contributes to Warfighter Protection), exposition of the FNC details in one setting before one audience and one panel allowed each participant to understand 1) the sub-capability inter-relations within each FNC and among the FNCs, and 2) how to best leverage and exploit these inter-relations for maximum aggregate FNC benefit.

For the remainder of the presentation week, the fifty-five Execution Managers presented their programs. The nominal presentation period was twenty minutes for actual presentation, ten minutes for questions and answers, and an additional five minutes for the reviewers to complete the evaluation forms. Some larger and more complex programs required more than twenty minutes, and smaller programs required less than twenty minutes.

Shortly after the review, the panel-averaged numerical results and integrative statistics were e-mailed to all the reviewers. The review managers then performed analyses and interpretations of the numerical results, and summarized the reviewers' comments in preparation for an Executive Session. These comment summaries were sent to the Executive Session audience shortly before the meeting; a summary of all the results was presented at the Executive Session. The final results and recommendations were used by senior DoN S&T management in the planning and budget allocation projections for the future DoN S&T program.

4. RESULTS OF REVIEW/ RECOMMENDATIONS

Because of the classified nature of the review, detailed results will not be presented. Instead, the types of results obtained, and the recommendations for action based on these results, will be outlined. Results were categorized into three types:

1) Overall 6.3 program results
2) Programs related to FNCs
3) Individual program results

4.1. Overall 6.3 Program Results

For the evaluation criteria Military Impact, Technical Approach, Program Execution, Transitionability, and Overall Item Evaluation, distribution functions of numbers of programs vs. rating bands (Low, Medium, High) were presented. No systemic overall 6.3 problems were uncovered.

4.2. Programs Related to FNCs

For the evaluation criterion Military Goal, the number of programs related to each FNC with strengths of relationships above parametrically-varied thresholds was obtained. In addition, the number of programs related to multiple FNCs was calculated. All 6.3 programs were related to at least one FNC with a strength of relationship of Medium or higher, and 95% of the 6.3 programs were related to at least one FNC with a strength of relationship of High. Some 6.3 programs were related to as many as eight FNCs with a strength of relationship of Medium or higher, and a few 6.3 programs were related to as many as four FNCs with a strength of relationship of High. Having this understanding of inter-relationships will be invaluable in helping the Execution Managers coordinate the program management and output among the IPTs.

The 6.3 programs were ranked by strength of relationship to each FNC. At the Executive Session, the principal S&T representative to each IPT discussed the potential role of the strongly related programs to addressing the FNC's goals.

4.3. Individual Program Results

The panel-averaged ratings for each 6.3 item for the six criteria were generated. These data were used to determine the aggregate relationships noted above. A regression analysis of the five component criteria against the Overall Item Evaluation criterion was performed, to determine which criteria had the most influence on bottom-line score (Overall Item Evaluation). Two criteria, Military Impact and Technical Approach, provided the bulk of the influence on the determination of bottom-line score. A model consisting of these two criteria predicted the bottom-line score to within two per cent. This is consistent with other large-scale reviews (DOE, 1982; Kostoff, 1997d).

This result should not be interpreted that the other three component evaluation criteria were unimportant. Rather, construction of a correlation matrix showed that the component criteria were strongly correlated, and the other three component criteria were subsumed under the two dominant criteria (Military Impact, Technical Approach).

For each of the fifty-five 6.3 items reviewed, a short description of the item's objectives and a summarization and integration of comments made by the Review Panel (categorized by the six review criteria) were generated. To arrive at these summary comments, the unabridged comments generated by the reviewers were read, and the main themes and messages were extracted. Where significant differences occurred between reviewers, minority and majority viewpoints were included.

4.4. Recommendations for Action

Numerical results were used to place the fifty-five 6.3 items in broad quality categories. Specific actions recommended for each item depended heavily on the comments from the reviewers, with special attention paid to the comments from the user/ customer representatives. In general, no corrective action was recommended for items that had good performance and execution, good transition potential, and strong relation to at least one FNC. Various levels of correction, including termination, were recommended for items that had the following characteristics:

- Insufficient commitment to transition
- "Core-Program" structure

- Insufficient FNC focus
- Insufficient demonstration focus

- Potential for high cost over-run

5. LESSONS LEARNED FROM REVIEW

There were many lessons learned from all phases of the 6.3 review, including the planning and consideration of alternative approaches, the conduct of the actual 6.3 review, and the post mortem analysis of the review's results and processes. Five of the major lessons will be described in this section. These lessons include:

1) value of performing a total S&T budget category review in one setting;
2) differences between 6.3 review and 6.1/ 6.2 reviews;
3) understanding effective use of information technology in program reviews;
4) value of adequate background material and review preparation, and
5) improving match between reviewer expertise and specific evaluation criteria requirements.

5.1. Value of Performing a Total S&T Budget Category Review in One Setting

There are two limiting cases by which an assemblage of programs can be reviewed. One method is to review the assemblage as a group, the other is to review the programs individually. Group reviews allow comparisons to be made across programs, but two compromises are necessary in real-world logistics-limited environments. Breadth is covered at the expense of depth, and the reviewer expertise per program will be smaller. Countering these compromises is the excellent normalization obtained with a single panel in a single setting. Individual reviews allow more in-depth assessment, and more specialty-focused reviewers. In addition, for a vertically-structured organization such as DoN S&T, individual program reviews (e.g., one 6.3 program) allow the other members of the vertical structure (e.g., related 6.1 and 6.2 programs) to be reviewed as well.

The typical DoN S&T review examines sub-groups of programs, usually spanning budget categories. The total 6.3 review showed that there was equal value in examining the total budget category at one setting, because of the comparative value. Selection of individual vs. group review of programs should depend on the overall review's objectives. An interspersing of both types of reviews over an organization's operational cycle is probably optimal. Neither approach is intrinsically superior.

5.2. Differences between 6.3 Review and 6.1/ 6.2 Reviews

Fundamentally, the objectives of reviewing 6.3 are not very different from those of reviewing 6.1 and 6.2. In both cases, military relevance and technical quality are the main drivers. However, while the 6.1 programs aim at achieving enhanced understanding of fundamental processes, the 6.3 programs aim at demonstrating products with desired affordability and performance characteristics. These differences tend to be reflected in the selection of specific criteria for each review type, in how the presentations address those criteria, and in the balance of types of reviewers selected for panel evaluations.

The 6.1 reviews focus on evaluating the advances in knowledge and the research questions answered, using criteria such as research merit, research approach, balance between experiment and theory, degree of innovation, and potential applications, while the 6.3 reviews use the criteria mentioned previously. The metrics have a different time scale involved. The 6.1 programs have a long-range focus; the 6.1 output metrics (papers, patents, etc) may have a short-term focus, but the 6.1 outcome metrics (benefit-cost ratio, rate of return, dollars saved, quality of life improvements) have a long-term focus. Many times, the 6.1 outcome metrics results can no longer be related to the research managers or performers or programs that they were designed to measure, and their operational utility can be called into question. For 6.3, the outcome metrics are much more closely related in time to the programs, managers, and performers these metrics were designed to measure, and a greater degree of accountability can be obtained from using the 6.3 outcome metrics.

While 6.1, 6.2, and 6.3 review panels all have S&T and customer/ user representation, the differences among panels tend to be in the relative emphasis of representation from the different communities. Across agencies, the 6.1 panels typically consist mainly of scientists and technologists, with some user/ customer representation, while the 6.3 panels typically have a much larger user/ customer fraction.

In those cases where 6.1 programs are reviewed with their 6.2 and 6.3 counterparts, as part of a larger vertical structure review (e.g., ONR's Department reviews), the panels tend to be relatively balanced with respect to community participation. These types of vertically-integrated structure reviews tend to be very informative, with substantial exchange of cross-category information. Any 'impedance mis-matches' across categories are easily detected, and corrections

can be readily recommended that will maximize vertical structure quality, as opposed to maximizing single category quality.

To repeat, single category and vertically-integrated structure reviews each have a unique role to play in an organization's overall strategic management process, and these roles depend on the review's specific objectives.

5.3. Understanding Effective Use of Information Technology in Program Reviews

One point became crystal clear in selecting appropriate information technology to support the review process. The following sequence should be obeyed religiously: Review objectives determine the metrics to be used; metrics determine the data to be gathered; metrics and data determine the types of reviewers selected; and metrics and data and reviewers jointly determine the process and supporting tools to be used. In particular, the Group-Ware selected should support the process and objectives, not drive them as is the all too familiar case in practice today. Furthermore, the Group-Ware needs to be specifically tailored to the process and objectives selected. The Group-Ware needs to be an integral component of the operational process, just as a particular scalpel serves as an integral component of a surgeon's repertoire. Efficient use of Group-Ware in the context of a network-centric review process (see Appendix 4) is discussed in Appendix 3.

5.4. Value of Adequate Background Material and Review Preparation

A major purpose of providing background material to all review participants before the presentations, especially to the review panel, is to insure that each participant will have a threshold level of understanding about each aspect of each program. A balance needs to be reached between the amount of material provided, and the amount that will be read by the reviewers. This balance will affect the structure of the material.

The 6.3 reviewers and audience were provided draft copies of the vugraphs to be presented at the actual review, about a week before the presentations. The vugraphs were posted on a password-protected Web site, and any other supportive material the presenters believed was important was added to the Web site as well. This background material proved adequate for the intended purpose. In other

program reviews, the first author has tended to provide two or three page narrative summaries for each program component to be presented. For example, if a $40 million Aircraft program review consists of presenting eight $5 million Aircraft component briefings (e.g., propulsion, aerodynamics, avionics), then the background material might consist of two or three page narrative summaries for each of the eight component areas, plus perhaps a three page summary of the total Aircraft program. This amount of background material is probably near the limit of what reviewers can be expected to read in traditional presentation-centered reviews, especially when their participation is pro bono, or near pro bono.

However, except for reviewers' time constraints, there appears to be no fundamental reason that much of the evaluation groundwork could not be done prior to the presentations. The Dutch STW (a government S&T sponsoring organization), for example, conducts one type of review entirely by mail (Van Den Beemt, 1991, 1997). If presentations are desired, and if sufficient programmatic material could be sent to the reviewers before the presentations, then much of the evaluation could be completed in advance of the presentations. Use of the new information technology, embedded in a facilitated process that encourages extensive interactions among reviewers and presenters, could enable this groundwork to be performed very efficiently, and not be overly burdensome on reviewers' time. One method for achieving this pre-presentation evaluation, based on experience gained with an innovation workshop (Kostoff, 1999a) and some experiences with other program reviews, is included in the description of a proposed network-centric review process (Appendix 4).

5.5. Improving Match between Reviewer Expertise and Specific Evaluation Criteria Requirements

In the 6.3 review, all the reviewers rated all the evaluation criteria. Yet some of the reviewers had substantial experience in technology development and less in military operations, whereas with other reviewers the converse was true. As a body, the reviewers covered all the evaluation criteria quite well with their aggregate expertise.

While the review results would probably be unchanged, it might be more efficient to have each reviewer's expertise matched more closely with each evaluation criterion. This can be accomplished in at least two ways. First, a weighting could be applied to each reviewer's rating for each evaluation criterion, based on the

reviewer's expertise relative to that criterion. Second, reviewers could be selected to rate specific criteria only.

The latter approach would probably be most desirable. Because of the large number of individuals that would be required as reviewers, implementation of such an approach has presented logistical difficulties in the past. Use of the new information technology, imbedded in a process that includes extensive interactions before the actual presentations (outlined above), would allow a much closer match between reviewers' expertise and specific evaluation criteria. It would allow the large number of reviewers required to achieve statistical significance for each criterion's ratings to be utilized efficiently.

One method of achieving this desirable match-up is included in the network-centric review process proposed in Appendix 4.

All the above lessons learned from the 6.3 review, lessons learned from other S&T reviews, and reasonable extrapolations therefrom, have been integrated into the proposed network-centric program review process described in Appendix 4. The key features of this network-centric S&T evaluation process are:

- Use of Group-Ware for real-time data entry and summary statistical displays
- Larger representation from technical communities due to logistics management with Group-Ware support
  - a) Use of many reviewers allows separation of Jury function (management decision-making) from Expert Witness function (technical judgment and testimony)
  - b) Use of many reviewers allows selection of reviewers with expertise in specific evaluation criterion for specific technical areas
- Expanded distribution of background material using Internet/ e-mail transmission
- Extensive e-mail interactions and preliminary evaluations before actual presentations
- Potential for completely remote reviews

6. SUMMARY AND CONCLUSIONS

A review of the total DoN S&T FY00 6.3 program was conducted by a senior DoN review panel. The review's purpose was to assess the 6.3 program from the perspectives of military relevance, technical quality, transitionability, and demonstration executability.

6.1. Evaluation Criteria

Five specific component criteria were used by the evaluation panel:

- Military Goal;
- Military Impact;
- Technical Approach/ Payoff;
- Program Executability; and
- Transitionability.

A sixth bottom-line criterion, Overall Item Evaluation, was also used

6.2. Evaluation Panel

The evaluation panel consisted of:

- ONR Executive Steering Committee;
- DoN S&T resource sponsor representatives;
- Marine Corps representatives;
- Advisors
  - 4a) FNC IPT Chairpersons
  - 4b) ONR Department Heads

6.3. Review Components

The major review components were:

1) Situation report presentations to the evaluation panel by the Chairpersons of the twelve FNC IPTs;
2) Technical presentations to the evaluation panel by the Execution Managers of the fifty-five 6.3 items;

3) Ratings and comments by the reviewers for each of the evaluation criteria for each 6.3 item
4) Processing of individual numerical entries to generate panel-averaged ratings, FNC distributions, and overall 6.3 program distributions; and
5) An Executive Session in which the numerical results were presented and placed in the larger FNC context.

6.4 Lessons Learned

Insights gained from both the planning and conduct of the review should be of considerable value when conducting future large-scale 6.3-type reviews, and include the following:

1) Provision of detailed programmatic descriptive material to the panelists and audience before the review is very useful; its value could be enhanced by e-mail interchange between the presenter or facilitator and the panelists before the presentations to clarify outstanding issues and allow for more effective use of actual meeting time.

2) Appropriate use of Group-Ware could allow
-Streamlining the review process with real-time data analysis and aggregation
-Remote reviewer participation, thereby minimizing travel and logistics problems
-More reviewers to participate in the process, producing a more representative sample of the technical community
-Reviewers to be selected for expertise in specific evaluation criteria only, thereby enhancing the credibility of each rating
-Sufficient expertise on the panel such that the Jury function (fully independent decision-making) can be separated from the Expert Witness (potentially conflicted technical judgment and testimony) function

3) When assessing quality of programs representing multiple disciplines, it is necessary to normalize. Evaluating all programs in one setting is an excellent way to accomplish this objective. Because of the realistic time constraints associated with a single-setting review, depth must be traded off for breadth. This trade-off is acceptable, as long as depth is evaluated by some means

during the S&T operational management cycle.

7. REFERENCES AND BIBLIOGRAPHY

Altura-BT, "Is Anonymous Peer-Review the Best Way to Review and Accept Manuscripts", MAGNESIUM AND TRACE ELEMENTS, 1990, Vol 9, Iss 3, pp 117-118.

Armstrong, J.S., "Why Conduct Journal Peer Review: Quality Control, Fairness, or Innovation", SCIENCE AND ENGINEERING ETHICS, 3:1, 1997.

ASTEC, "Funding the Fabric - Should Commonwealth Government Competitive Research Granting Schemes Contribute More to Research Infrastructure Costs?", Canberra: Australian Government Publishing Service, 1991.

Brown, E. A., "Conforming the Government R&D Function with the Requirements of the Government Performance and Results Act: Planning the Unplannable? Measuring the Unmeasurable?", SCIENTOMETRICS, 36:3, 1996.

Buechner, Q., "Proposal Costs," JOURNAL OF THE SOCIETY OF RESEARCH ADMINISTRATORS, 5, 47-50, 1974.

Ceci, S. J. & Peters, D., "How Blind Is Blind Review?," AMERICAN PSYCHOLOGIST, 39 (2), 1491-1494, 1984.

Chubin, D. E. and Hackett, E. J., "Peerless Science: Peer Review and U. S. Science Policy", State University of New York Press, Albany, NY, 1990.

Clayson-DB, "Anonymity in Peer-Review - Time for a Change - Comment", REGULATORY TOXICOLOGY AND PHARMACOLOGY, 1995, Vol 22, Iss 1, pp 101-101

Cox-D Gleser-L Perlman-M Reid-N Roeder-K, "Report of the AD-Hoc-Committee-on-Double-Blind-Refereeing", STATISTICAL SCIENCE, 1993, Vol 8, Iss 3, pp 310-317

Delcomyn-F, "Peer-Review - Explicit Criteria and Training Can Help", BEHAVIORAL AND BRAIN SCIENCES, 1991, Vol 14, Iss 1, pp 144-144

DOE, "An Assessment of the Basic Energy Sciences Program", Office of Energy Research, Office of Program Analysis, Report No. DOE/ER-0123, March 1982.

DOE, "Evaluation of Alternate Magnetic Fusion Concepts 1977", DOE/ET-0047, U.S. Department of Energy, Assistant Secretary for Energy Technology, Office of Fusion Energy, May 1978.

Fielder-JH, "Disposable Doctors - Incentives to Abuse Physician Peer-Review", JOURNAL OF CLINICAL ETHICS, 1995, Vol 6, Iss 4, pp 327-332

Goodstein-D, "Ethics and Peer-Review - Commentary", STEM CELLS, 1995, Vol 13, Iss 5, pp 574-574

GPRA, Government Performance and Results Act of 1993 (PL 103-62), 1993

Greengrass, E, "Information Retrieval: an Overview", National Security Agency, Report Number TR-R52-02-96, 28 February 1997.

Gresty-MA, "Peer-Review and Anonymity", NEURO-OPHTHALMOLOGY, 1995, Vol 15, Iss 6, pp 281-282

Gupta-VK, "Should Intellectual Property Be Disseminated by Forwarding Rejected Letters Without Permission", JOURNAL OF MEDICAL ETHICS, 1996, Vol 22, Iss 4, pp 243-244

Hall, D., and Nauda, A., "An Interactive Approach for Selecting IR&D Projects," IEEE TRANSACTIONS ON ENGINEERING MANAGEMENT, Vol. 37, No. 2, May 1990.

Hensley, O., Gulley, B., and Eddleman, J., "Evaluating Development Costs for a Proposal to a Federal Agency," JOURNAL OF THE SOCIETY OF RESEARCH ADMINISTRATORS, 12, 35-39, 1980.

Keown-D, "The Journal of Buddhist Ethics - An Online Journal", LEARNED PUBLISHING, 1996, Vol 9, Iss 3, pp 141-145

Kostoff, R. N., "GPRA Science and Technology Peer Review", SciCentral, www.scicentral.com, 1998.

Kostoff, R. N., "Peer Review: The Appropriate GPRA Metric for Research", SCIENCE, Volume 277, 1 August 1997a.

Kostoff, R. N., "Research Program Peer Review: Principles, Practices, Protocols", http://www.dtic.mil/dtic/kostoff/index.html, 1997b.

Kostoff, R. N., "The Principles and Practices of Peer Review", in: Stamps, A. E., (ed.), SCIENCE AND ENGINEERING ETHICS, Special Issue on Peer Review, 3:1, 1997c.

Kostoff, R. N., "The Handbook of Research Impact Assessment," http://www.dtic.mil/dtic/kostoff/index.html, 1997d. Also, DTIC Technical Report Number ADA296021.

Kostoff, R. N., Science and Technology Innovation, TECHNOVATION. 19:10, 1999a. Also, http://www.dtic.mil/dtic/kostoff/index.html

Kostoff, R. N., Eberhart, H. J., and Toothman, D. R. "Hypersonic and Supersonic Flow Roadmaps Using Bibliometrics and Database Tomography". JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE. 15 April, 1999b.

Laband, D. N., "A Citation Analysis of the Impact of Blinded Peer-Review", JAMA, 272:2, 1994.

Moran-G, "Ethical Questions About Peer-Review", JOURNAL OF MEDICAL ETHICS, 1992, Vol 18, Iss 3, pp 160-160

Neetens-A, "Should Peer Reviewers Shed the Mask of Anonymity", NEURO-OPHTHALMOLOGY, 1995, Vol 15, Iss 3, pp 109-109.

NRC. "Peer Review in Environmental Technology Development Programs". National Academy Press. Washington, DC. 1998.

NRC. Evaluating Federal Research Programs: Research and the Government Performance and Results Act. National Academy Press. Washington, DC. 1999.

NRC. Peer Review in Environmental Technology Development Programs. National Academy Press. Washington, DC. 1998.

Nylenna-M Riis-P Karlsson-Y, "Multiple Blinded Reviews of the 2 Manuscripts - Effects of Referee Characteristics and Publication Language", JAMA-JOURNAL OF THE AMERICAN MEDICAL ASSOCIATION, 1994, Vol 272, Iss 2, pp 149-151

Sutherland-HJ Meslin-EM Dacunha-R Till-JE, "Judging Clinical Research Questions - What Criteria Are Used", SOCIAL SCIENCE & MEDICINE, 1993, Vol 37, Iss 12, pp 1427-1430

USNRC, "Peer Review for High-Level Nuclear Waste Repositories: Generic Technical Position". NUREG-1297. Washington, DC. U.S. Nuclear Regulatory Commission, 1988.

Van den Beemt, F.C.H.D, "The Right Mix: Review by Peers as Well as by Highly Qualified Persons (Non-Peers)", Australian Research Council Commissioned Report: "Peer Review Process" No.54 (1997), 153-164.

Van den Beemt, F.C.H.D. and Le Pair, C., "Grading the Grain: Consistent Evaluation of Research Proposals", RESEARCH EVALUATION, 1:1, 1991.

Weinberg, A. M., "Criteria for Evaluation, a Generation Later", In The Evaluation of Scientific Research (ed. Ciba Foundation), pp. 3-12.Chichester: John Wiley & Sons, 1989.

8. APPENDICES

APPENDIX 1 - PRINCIPLES OF HIGH QUALITY PEER REVIEW

This appendix presents, in priority order, the underlying principles necessary for high quality peer and internal reviews, and updates and expands the principles contained in Kostoff (1997b). While the paper is targeted toward program peer and internal review, most of the principles are applicable to multiple types of peer and internal reviews (proposals, programs, procedures, manuscripts, faculty or dissertations). The first author's experience, based on examining the peer review literature, conducting many peer review experiments (e.g., Kostoff, 1988), and managing hundreds of peer reviews, leads to the following conclusions about the factors critical to high-quality peer review (Kostoff, 1995, 1997a, 2001b).

PRINCIPLES:

1) Senior Management Commitment
The most important factor in a high-quality S&T evaluation is the serious commitment to high-quality S&T evaluations of the evaluating organization's most senior management with evaluation decision authority, and the associated emplacement of rewards and incentives to encourage such evaluations. Incorporated in senior management's commitment to quality evaluations is the assurance that a credible need for the evaluation exists, as well as a strong desire that the evaluation be structured to address that need as directly and completely as possible.

2) Evaluation Manager Motivation
The second most important factor is the operational evaluation manager's motivation to perform a technically credible evaluation. The manager:

a) sets the boundary conditions and constraints on the evaluation's scope;

b) selects the final specific evaluation techniques used;

c) selects the methodologies for how these techniques will be combined/ integrated/ interpreted, and

d) selects the experts who will perform the interpretation of the data output from these techniques.

In particular, if the evaluation manager does not follow, either consciously or subconsciously, the highest standards in selecting these experts, the evaluation's final conclusions could be substantially determined even before the evaluation process begins. Experts are required for all the evaluation processes considered (peer review, retrospective studies, metrics, economic studies, roadmaps, data mining, and text mining), and this conclusion about expert selection transcends any of these specific applications.

3) The third most important factor is the transmission of a clear and unambiguous statement of the review's objectives (and conduct) and potential impact/ consequences to all participants. This statement should occur at the very beginning of the review process.

4) Competency of Technical Evaluators
The fourth most important factor is the role, objectivity, and competency of technical experts in any S&T evaluation. While the requirements for experts in peer review, retrospective studies, roadmaps, and text mining are somewhat obvious, there are equally compelling reasons for using experts in metrics-based evaluations. Metrics should not be used as a stand-alone diagnostic instrument (Kostoff, 1997b). Analogous to a medical exam, even quantitative metrics results from suites of instruments require expert interpretation to be placed into proper context and gain credibility. The metrics results should contribute, and be subordinate, to an effective peer review of the technical area being examined.

Thus, this fourth critical factor consists of the evaluation experts' competence and objectivity. Each expert should be technically competent in his subject area, and the competence of the total evaluation team should cover the multiple S&T areas critically related to the science or technology area of present interest. In addition, the team's focus should not be limited to disciplines related only to the present technology area (that tends to reinforce the status quo and provide conclusions along very narrow lines). It should be broadened to disciplines and technologies that have the potential to impact the overall evaluation's highest-level

objectives (that would be more likely to provide equitable consideration to revolutionary new paradigms).

5) Selection of Evaluation Criteria

The fifth most important factor is selection of evaluation criteria (Delcomyn, 1991; Sutherland, 1993; Weinberg, 1989). These criteria will depend on the:

- interests of the audience for the evaluation,
- nature of the benefits and impacts,
- availability and quality of the underlying data,
- accuracy and quality of results desired,
- complementary criteria available and suites of diagnostic techniques desired for the complete analysis,
- status of algorithms and analysis techniques, and
- capabilities of the evaluation team.

For evaluating basic research proposals, the three main criteria are research merit, research approach, and team quality (DOE, 1982; Kostoff, 1992, 1997a). For research sponsored by a mission-oriented organization, a fourth criterion related to mission relevance is useful. To ensure that this mission relevance criterion does not filter out the more basic research oriented proposals, a very liberal interpretation of mission relevance is necessary. For basic research, a nearer-term relevance criterion, such as transition or utility, correlates better with overall proposal quality score than does a longer-term criterion (Kostoff, 1992). Use of a fifth criterion for overall research quality is essential, and makes it possible to incorporate the effects of unlisted criteria that the reviewer feels is important for considering a specific proposal. For example, reviewers might feel that an agency proposal is more appropriate for sponsorship by industry than by government. In this case, the proposal could receive a low overall rating, even though the listed component technical criteria were rated very high.

6) Relevance of Evaluation Criteria to Future Action

A factor of equal importance to evaluation criteria selection is one that has been violated in almost every metrics briefing the author has attended spanning many government agencies, industrial organizations, and academic institutions. In general, this factor tends to be violated for the evaluation criteria used in any of

the evaluation approaches under the decision aids umbrella. The factor will be stated in terms of a metrics-based evaluation, but it should be considered as applicable to all evaluation techniques.

EVERY S&T METRIC, AND ASSOCIATED DATA, PRESENTED IN A STUDY OR BRIEFING SHOULD HAVE A DECISION FOCUS; IT SHOULD CONTRIBUTE TO THE ANSWER OF A QUESTION WHICH IN TURN WOULD BE THE BASIS OF A RECOMMENDATION FOR FUTURE ACTION.

Metrics and associated data that do not perform this function become an end in themselves, offer no insight to the central focus of the study or briefing, and provide no contribution to decision-making. They dilute the theme of the study, and, over time, tend to devalue the worth of metrics in credible S&T evaluations. Because of:

1) the political popularity and subsequent proliferation of S&T metrics;
2) the widespread availability of data; and
3) the ease with which this data can be electronically gathered/ aggregated/ displayed,

most S&T metrics briefings and studies are immersed in data geared to impress rather than inform. While metrics studies provide the most obvious examples, this conclusion can be easily generalized to any of the evaluation methods.

7) Reliability of Evaluation

Another factor of equal importance is reliability or repeatability. To what degree would an S&T evaluation be replicated if a completely different team were involved in selection, analysis, and interpretation of the basic data? If each evaluation team were to generate different evaluation criteria, and in particular, generate far different interpretations of these criteria for the same topic, then what meaning or credibility or value can be assigned to any S&T evaluation (Cole, 1981)? To minimize repeatability problems, a diverse and representative segment of the overall competent technical community should be involved in the construction and execution of the evaluation.

8) Evaluation Integration

A fourth factor of equal importance is the seamless integration of evaluation processes in general into the organization's business operations. Evaluation processes should not be incorporated in the management tools as an afterthought, as is the case in practice today, but should be part of the organization's front-end design. This allows optimal matching between data generating/ gathering and evaluation requirements, not the present procedure of force fitting evaluation criteria and processes to whatever data is produced from non-evaluation requirements.

9) Global Data Awareness

A fifth factor of equal importance is data awareness (Kostoff, 1999a, 2000a, 2000b, 2001d, 2003a, 2003c). In all of the decision aids, placement of the technology of interest in the larger context of technology development and availability world-wide is an absolute necessity. This tends to be a central deficiency of most management decision aids. Lack of S&T documentation, inaccessibility of S&T that is documented, inability to retrieve S&T documents due to poor retrieval methods, inability to extract information from large retrievals, and general lack of interest and will in global data awareness, mitigate against attaining comprehensive global data awareness.

10) Normalization across Technical Disciplines

For evaluations that will be used as a basis for comparison of S&T programs or projects, the next most important factor is normalization and standardization across different S&T areas. For S&T areas that have some similarity, use of common experts (on the evaluation teams) with broad backgrounds that overlap the disciplines can provide some degree of standardization (Kostoff, 1988, 1997a). For very disparate S&T areas, some allowances need to be made for the relative strategic value of each discipline to the organization, and arbitrary corrections applied for benefit estimation differences and biases. Even in this case of disparate disciplines, some normalization is possible by having some common team members with broad backgrounds contributing to the evaluations for diverse programs and projects (Van den Beemt, 1997). However, normalization of the criteria interpretation for each science or technology area's unique characteristics is a fundamental requirement. Because credible normalization requires substantial time and judgment, it tends to be an operational area where quality is sacrificed for expediency.

11) Reviewer Anonymity

A factor of equal importance to normalization is secrecy: reviewer anonymity and reviewee non-anonymity (Altura, 1990; Clayson, 1995; Gresty, 1995; Neetens, 1995). If honest and frank viewpoints on the intrinsic quality of the research under review are desired, the reviewer must remain anonymous to all but the review manager. Rewards are few for a reviewer making strong negative statements about a proposal (or research paper or program), and resulting retributions and resentments to the reviewer may far outweigh the intrinsic benefits to science of honest and forthright judgment statements.

"Blind reviewing," the withholding of the reviewee's name and affiliation from the reviewer, has been used for the noble purposes of providing fairer reviews of work by unknown researchers or by researchers from less prestigious institutions, and to eliminate bias based on personal characteristics such as gender (Ceci, 1984; Laband, 1994; Cox, 1993; Nylenna, 1994). However, studies of proposed and existing research evaluations have shown that team quality was the most important variable in determining overall project quality (DOE, 1982). Removing the identity of the reviewee from the research under review is akin to solving an equation after eliminating the dominant term. As a result, rather than eliminate the key variable of researcher identity, it may be more important to select additional reviewers who will broaden the review group's perspective and address the "right job" aspects of the research project. This will help insure that outmoded, albeit frequently cited, research is not promulgated in perpetuity, and that fresh perspectives of new paradigms will receive the attention they deserve.

12) Cost of S&T Evaluations

The next critical factor for quality S&T evaluations is cost (ASTEC, 1991; Buechner, 1974; Hensley, 1980; Kostoff, 1995, 1997a). The true total costs of peer review can be considerable, but tend to be ignored or understated in most reported cases. For high quality peer reviews, where sufficient expertise is represented on the review group, total real costs will dominate direct costs (Kostoff, 1995, 1997a). The major contributor to total costs is the time of all the individuals involved in executing the review, including staff, reviewer, and presenter time. If a substantial audience is in attendance, then audience time should be included in review costs. With high quality performers and reviewers,

time costs are high, and the total review costs can be non-negligible. For sponsor environments where a large number of proposals are rejected, and where multiple proposals to different sponsors are the norm, peer review costs per funded proposal increase dramatically in proportion to the ratio of proposals reviewed to proposals funded. Accurate cost analyses should not be neglected in designing a high quality proposal, manuscript, or program peer-review process.

13) Maintenance of High Ethical Standards

The final critical factor, and perhaps the foundational factor, in any high quality S&T evaluation is the maintenance of high ethical standards throughout the process. There is a plethora of potential ethical issues (Fielder, 1995; Goodstein, 1995; Gupta, 1996; Keown, 1996; Moran, 1992), including technical fraud, technical misconduct, betraying confidential information, and unduly profiting from access to privileged information. This stems from an inherent bias/ conflict of interest in the process when real experts are desired to participate in every aspect of an S&T evaluation. The evaluation managers need to be vigilant for undue signs of distortion aimed at personal gain.

APPENDIX 2 - EVALUATION CRITERIA USED IN 6.3 REVIEW

Evaluator Name: Date: Monday – 2 August

Evaluator Organization: Time: 1345

S&T 6.3 Thrust/ATD/MDD Program Title: Advanced Multi-Function RF System

1) MILITARY GOAL (Enter ONE INTEGER between 1 and 10 for each FNC)

HI			MED				LO		
10	9	8	7	6	5	4	3	2	1

FNC		FNC	
Information Distribution		Missile Defense	
Time Critical Strike		Platform Protection	
Decision Support Systems		Expeditionary Logistics	
Autonomous Operations		Warfighter Protection	
Littoral ASW		Capable Manpower	
Total Ownership Cost Reduction		Organic MCM	

(Circle ONLY ONE number for each criterion)

	HI			MED			LO			
1. MILITARY IMPACT	10	9	8	7	6	5	4	3	2	1
2. TECHNICAL APPROACH	10	9	8	7	6	5	4	3	2	1
3. PROGRAM EXECUTABILITY	10	9	8	7	6	5	4	3	2	1
4. TRANSITIONABILITY	10	9	8	7	6	5	4	3	2	1

5. OVERALL ITEM EVALUATION	10	9	8	7	6	5	4	3	2	1

Comments:

6.3 Review Scoring Definitions and Values

1) MILITARY GOAL
How important is the Thrust's 6.3 component or the ATD/Maritime Defense Demonstration to the designated Future Naval Capabilities?

HI - Critical to one or more of the 12 designated Future Naval Capabilities
MED - Addresses one or more of the 12 designated Future Naval Capabilities
LO - Does not address one of the 12 designated Future Naval Capabilities

2) MILITARY IMPACT
What is the Thrust's 6.3 component or ATD/Maritime Defense Demonstration's potential for military capability improvement? What are the products?

HI - Revolutionary
MED - Substantial
LO - Incremental

3) TECHNICAL APPROACH
Why was this approach taken?

HI - Better technical payoff than alternate approaches
MED - Equivalent technical payoff to alternate approaches
LO - Worse technical payoff than alternate approaches

4) PROGRAM EXECUTABILITY
What is the probability that the Thrust's 6.3 component or ATD/Maritime Defense Demonstration's technical targets can be demonstrated at the stated costs and schedule?

HI – Near certainty
MED – Probably
LO – Unlikely

5) TRANSITIONABILITY
What is the probability that the Thrust's 6.3 component or ATD/Maritime Defense Demonstration will result in transition to higher category development or acquisition if successful?

HI – Solid financial commitment by transitionee
MED – Solid support without financial commitment by transitionee
LO – No support (including negative support) by transitionee

6) OVERALL ITEM EVALUATION
What is the bottom-line Thrust's 6.3 component or ATD/Maritime Defense Demonstration's quality score, based on evaluation criteria above and any other criteria deemed important by reviewers?

HI - Revolutionary improvements in military and technology capabilities
MED - Substantial improvements in military and technology capabilities
LO - Incremental improvements in military and technology capabilities

APPENDIX 3 - INTEGRATED GROUP-WARE FOR PROGRAM PEER REVIEW

A3-1) Group-Ware Software System

The main intention in using groupware was to allow electronic collection of data, ratings and comments, that could be used for immediate analysis, documentation, and display. Two groupware systems were considered in preparation for the 6.3 Program Review – the first option was commercially available (Ventana System's Group Systems), whereas the second was developed in-house. Time constraints lead to the use of a hybrid of the two systems.

The commercial groupware system used at the 6.3 Program Review is a proven software, typically used in a voting / rating scenario. The software was networked to several computers, that allowed data entry personnel to input data simultaneously. It also allowed for real-time compilation of data, including basic analysis such as calculated mean values, distribution functions of the ratings, standard deviations, and histogram plots of the voting results. Drawbacks in this groupware system included the limited types of output, and incompatibility with other commercial softwares such as Microsoft (MS) Excel or MS Powerpoint. Output files had to be manipulated by experts to allow further analyses not performed by the groupware system.

A groupware simulating database systems was developed as an alternative. This approach was later tested, and proved to be far more powerful than the commercial system for the specific application due to its flexibility. The groupware system used readily available and internally compatible software (Microsoft ACCESS, Excel, PowerPoint). The database approach could be tailored for any review scenario requiring electronic data collection and instantaneous analysis, documentation, and display. This system could be pre-programmed with user defined requirements, such that only desired / specific outputs or analyses are performed. Outputs could be manipulated in various ways (filtering, sorting, variety of plots, etc.). Numerical ratings and text comments could be automatically documented in a presentable pre-formatted report. Outputs are fully compatible with all word processing and spreadsheet software packages.

One of the premiere features of the developed database system is the ability to develop and tailor graphical user interfaces (GUI), with simple icons to facilitate data entry, and thereby reduce the probability of error. GUIs can also be programmed such that the user can navigate through the program and retrieve and display the desired outputs. This system is now available for use by the FNC IPTs for decision-making processes, or by other users for DoN S&T reviews.

APPENDIX 4 – NETWORK-CENTRIC PEER REVIEW

I) INTRODUCTION

The objective of the proposed network-centric peer review is to evaluate a large ongoing S&T program, using a representative segment of the technical community, and employing whatever information technology is required to substantially enhance the quality of the review. Network-centric peer review uses the power of modern communication networks and information technology to expand greatly the number of people that can participate in real-time peer reviews, and expands greatly the access to data that can support all aspects of peer review. This technology allows diverse review operational modes such as the Science Court to be considered seriously, and allows the jury function of peer review to be independent from the higher conflict potential expert reviewer/ witness function The operational architecture required for network-centric peer review may differ little from the architecture required for its parent network-centric strategic management. Since all strategic management components need to be integrated for optimal synergistic benefits, implementation of network-centric peer review should occur in parallel with implementation of the other components of network-centric strategic management.

This appendix addresses:

*information technology advances and their potential impact on peer review;
*an implementation procedure for a network-centric peer review process;
*research opportunities for network-centric peer review.

II) INFORMATION TECHNOLOGY ADVANCES

In recent years, advances in computer hardware have resulted in much higher computational speed systems with massive amounts ofrapidly-accessible storage space. In parallel with the hardware advances are software improvements that allow organization and 'mining' of the transmitted data, and architecture implementations that allow large networks of disparate data sources (whether sensors, humans, structured databases, or other types) to be linked. With such

network architectures readily available, one person can communicate with many individuals at once, and the input from many individuals and data sources can be collected, integrated, and analyzed in real time. The implications for peer review in particular, and for strategic management in general, are enormous. One of the major (justified) criticisms of peer review (and of road-maps, metrics, data mining, information retrieval, S&T planning, S&T evaluation, S&T transitioning, and other strategic management decision support aids) has been that only a small fraction of the relevant communities and available data are being accessed when these decision aids are being exercised. Logistics costs and time delays have limited the magnitude of information and people available to contribute to these decision aids' outputs, especially when time frames approximating real-time are required. Now, the hardware and software in combination with the network architectures, and especially supported by *individuals who understand the relation between the information technology capabilities and the decision aid requirements*, allow these logistics-based limitations to be removed.

III - POTENTIAL IMPACT OF INFORMATION TECHNOLOGY ADVANCES ON PEER REVIEW

First, the potential impact of information technology advances on the different temporal segments of peer review will be estimated. Then, the potential impact of information technology advances on the different quality principles will be discussed. In the following section, these concepts and estimates will be crystallized and integrated into a proposed network-centric review process.

III -1) Impact on Temporal Segments

This discussion will be based on the assumption that one component of a research program peer review will be a meeting that some, not necessarily all, of the participants will attend. Conduct of a meeting-based research program peer review can be categorized into three stages: a pre-meeting phase, the actual meeting, and a post-meeting phase.

III - 1 - A) Pre-Meeting Phase

The main goal of the pre-meeting phase is to inform and prepare all the participants sufficiently that little time is wasted during the actual meeting phase. Standard peer reviews today allow the various review participants to receive

summary background material, to be read by the time of the meeting. An interdisciplinary workshop conducted by the author in December 1997 (Kostoff, 1999a) went one step further. Participants exchanged ideas by e-mail, and all participants were involved in each e-mail. By the time of the meeting, many of the issues had been greatly clarified. However, what could be envisioned in this pre-meeting phase if network-centric peer review were operable, utilizing much of the power of available information technology?

First, a substantially larger amount of data could be made accessible to each review participant, since the network could be structured to allow each node (participant) ready access to every other node (data source/ participant). Second, a substantially larger number of participants could be involved in the review, limited only by the extent of the network architecture. Third, a real time iterative rating, learning, and subsequent presentation modification process could be established. New concepts could be dialogued and improved, presentations could be critiqued and rated preliminarily, and greatly modified for the meeting. Some types of reviews could be conducted entirely without physical presence, whereas those that required an actual meeting would have most of the time-delaying issues examined beforehand. In summary, this phase could accommodate substantially more data and participants than at present, could integrate and analyze this data in real-time, and could provide feedback in a continuous short-turnaround mode. It could also provide a period of reflection and gestation, as concepts became more integrated with the passage of time. How could this network-centric pre-meeting phase be envisioned to affect the next actual meeting phase?

III - 1 - B) Meeting Phase

First, the actual review panel could consist of hundreds or more of experts, some of whom are on-site and the remainder are off-site. All would be linked through the network architecture, and the off-site participants may be video-tele-conferenced to the presentation material as well. These features allow the review process to be decentralized, either partially or fully, and provide much greater flexibility in time and location scheduling. They also allow a greater diversity of reviewers to be used, in technical areas ranging from closely aligned with the focused presentation themes to very disparate disciplines that could contribute innovative insights to the target themes and offer the possibility of real breakthroughs.

All data input would be mechanized, and instantly recorded. Statistical analyses could be performed on the data, at the level of each presentation and integrated over all presentations. This integrative analysis would show how each project's ratings would influence overall rankings and overall parametric criteria, thus placing local decisions in their global context. All the background data, the reviewers' ratings and comments, and other supportive data, would be available instantly to all participants. This latter feature would allow real-time Delphi processes, or modifications of comments and ratings, to be conducted at the end of the presentation period, or in dedicated Executive Sessions. The availability of large amounts of data of all types and large numbers of experts in diverse areas might allow the addition of extra evaluation criteria to be employed usefully, and offer additional perspectives on the S&T being reviewed. What impact could a network-centric meeting process have on the final post-meeting phase?

III - 1 - C) Post-Meeting Phase

The post-meeting phase would have some analogies to the pre-meeting phase, with more focus on integration of new concepts and identification of solutions/ modifications to problem areas identified, stimulated by the intense interactions from the highly efficient meeting phase. Final rankings, comments, and decisions would be obtained iteratively with the availability of the integrated comments and statistics, and a comprehensive integrated report could be assembled from the diverse reviewers effortlessly.

III - 2) Impact on Principles of High Quality

III - 2 - A) Need for Synergy and Integration

In the preface to the high quality principles section, the main theme expounded was that peer review, and the complementary decision aids as well, needed to be an integral component of the overall strategic management process. If peer review, or any of these decision aids, are treated as add-ons or independent entities, the power of these techniques and value to the sponsoring organization are diminished substantially. These techniques are interlocking, their operation is symbiotic, and their benefits are synergistic. For network-centric peer review to achieve its full potential, it must be integrated fully into the network-centric strategic management process. Thus, the requirements for successful operation

of network-centric peer review are more severe than for traditional peer review, because the operational targets and potential roadblocks are at a higher level.

For example, if data mining is not performed using all the global data sources available as well as the human and computer analytic and interpretive capabilities, then a gap will exist in the data available for comparing programs under review with the state-of-the-art. This in turn will affect the use of metrics to gauge the comparisons, and road-maps to show project and technology linkages. The impact of data-deficient peer review on strategic planning will result in greater uncertainty in the planning process and products, and will be translated into greater uncertainty in the project selection, management, and transition processes and products.

Thus, a full-scale network-centric strategic management process must eventually be developed, of which the peer review component is one element. However, once the architecture has been established for a network that links the S&T performer/ management/ oversight/ acquisition/ operational/ vendor communities, then

- peer review can be accomplished readily in the network-centric mode,
- road-maps can be easily generated in the network-centric mode,
- planning can be performed efficiently in a network-centric mode,
- multi-discipline multi-category multi-performer multi-user programs can be coordinated and managed effectively in the network-centric mode,
- Integrated Product Teams can conduct planning and operations in a highly decentralized network-centric mode, and
- even marketing and sales can be conducted in a network-centric mode using all the resources of organizations/ nations/ and international communities.

The key point here is that it is the architectural structure, and the inherent logic that links the nodes of the network, that are central to the effective operation of all these seemingly diverse components of strategic management. Once the architecture has been constructed, and the data control established, successful operation of the strategic management tactical elements ceases to be a critical path item.

III - 2 - B) Impact on Specific Principles

The first three principles of high quality peer review listed in Appendix 1 focus on management commitment, incentives, motivation, and statement of objectives. These provide a context, or set the stage, for conducting a high quality peer review, but would not be impacted by the specific tools employed during the review.

The fourth principle, Evaluator Competency, could be impacted substantially by network-centric operation. Three of the critiques related to evaluator competency in peer reviews are:

- that not all technical areas are covered adequately by relatively small panels used in peer reviews,
- even in those covered areas, the sample of the community is too small to be representative, and
- there are many facets of related technical and non-technical areas that the panel does not cover as a body because of the narrow technical focus.

Network-centric operation would allow many representatives from any technical speciality of interest, representatives from all technical areas involved, and representatives from areas that go beyond the purely technical (users of the technology, impactees, environmental, regulatory, etc.). Because time commitments of reviewers would be reduced due to less need for travel, and because high quality reviewers tend to be busy time-restricted people, more high quality reviewers would be available to participate in the review process, further raising the quality level of the review.

There is another potential benefit related to the Evaluator Competency criterion that deals with the evaluators' operational mode. In the vast majority of traditional S&T peer reviews, the panel has a dual role/ function. It serves as (hopefully) an impartial jury, and serves as an expert witness/ reviewer body as well. This is intrinsically different from the legal system, where the jury and the witnesses/ experts are separate bodies, with separate responsibilities and separate individual requirements. Combining the jury and witness/ expert functions has the potential for serious conflict. The combination problem arises mainly due to the finite

panel size, and the logistical inability to handle large numbers of witnesses/ experts in parallel with panel operation.

There have been attempts to conduct peer reviews in which the jury function is executed by one group, and the expert/ witness function is executed by an entirely distinct group (DOE, 1978; Van den Beemt, 1997). The Science Court procedure used by the first author to evaluate competing alternate magnetic fusion concepts is one example (DOE, 1978; Kostoff, 1997d). The first author's experience with the Science Court was that it was a very valuable process, but very time consuming and unwieldy. Network-centric operation would convert the Science Court into a much more manageable and powerful process.

Thus, network-centric operation offers potential benefits in either panel mode of operation. In the case where the panel operates as both the jury and expert/ witness body, network-centric operation expands the number of participants to insure expertise coverage of all criteria. In the case where the jury and witness/ expert body are separate, network-centric operation still insures expert coverage of all criteria, but allows the panel to function as a relatively independent conflict-free jury.

The next principle that could be affected by network-centric operation is Evaluation Criteria. With the expanded access to data allowed by network-centric operation, criteria could be added for which data could be obtained straight-forwardly. For example, suppose knowledge of specific types of impact was an important criterion, but the data by which impact would be evaluated were not readily available. Under traditional peer review, that criterion might not be used, but under network-centric operation, that criterion could be employed due to ready data availability on impact.

The criterion of Reliability would be impacted substantially by network-centric operation. With a large sample from the relevant communities, degree of representativeness is no longer an issue, and the repeatability of the results over different panels becomes a moot point. In addition, much more data becomes available for incorporation into the evaluation, and statistical representativeness effectively disappears as a data issue.

The Data Awareness criterion would obviously be affected to a large extent. Network-centric operation allows massive amounts of global data to be accessed, filtered, mined, interpreted, and evaluated. Bibliometric analysis capabilities will allow the performers, institutions, and countries that are sponsoring/ performing S&T to be identified, thereby enhancing the potential for leveraging and exploitation, and minimizing the opportunities for excessive redundancy. Along with limited numbers of reviewers, limited access to data is a major deficiency of present day peer reviews that would be overcome by network-centric operation.

The Secrecy criterion could be impacted to some degree. Network-centric operation could allow people at remote sites to participate as reviewers/ expert witnesses without their identity being revealed to other participants in the process. This enhanced anonymity would allow for greater open-ness and frank-ness, ultimately yielding a more useful product.

The Cost criterion would be impacted, due to the reduced travel requirement, and the reduced facilities requirement. Since time commitments would be reduced as well, high caliber typically busy people would be more likely to serve, and a higher quality product would also result concomitant with the lower cost.

IV - IMPLEMENTATION OF A NETWORK-CENTRIC REVIEW PROCESS

IV - 1) Background

The first author has conducted meetings/ reviews that have made some use of network capabilities. These include the review of the Department of the Navy's total Advanced Technology Development program described in the text, and an innovation workshop on Autonomous Flying Systems. The lessons learned from conducting these meetings/ reviews will be integrated with the principles of high quality peer review in Appendix 1 and the network concepts of this appendix to outline an operational implementation for a high quality network-centric S&T program peer review.

The objective of the review is to evaluate a large ongoing S&T program, using a representative segment of the technical community, and employing whatever information technology is required to substantially enhance the quality of the

review. For illustrative purposes only, the parameters of the Department of the Navy Advanced Technology Development program review described in the main text will be used in the following discussion.

IV - 2) Definition of Evaluation Criteria

In the proposed network-centric review, after the objectives and goals have been specified, the first operational step would be to define the evaluation criteria. These are the metrics that would allow quantitative determination of progress toward the goals and objectives. For mission-oriented organizations, there tend to be two over-arching evaluation criteria: mission-relevance and technical quality. For a variety of reasons, including the analysis of progress in achieving sub-goals and objectives, additional supportive criteria tend to be employed in reviews. For the proposed review, assume the same criteria are used as were employed in the Department of the Navy illustrative example: Military Goal; Military Impact; Technical Approach/ Payoff; Program Executability; and Transitionability. In combination, these criteria will help answer the question: Will this program result in a high impact high-quality militarily relevant product with high probability of meeting cost, schedule, and performance targets?

IV - 3) Selection of Review Taxonomy

The second operational step is selection of a taxonomy for the review. *A cardinal rule in assessment is that a program should be reviewed using the same taxonomy by which it was selected and managed.* Otherwise, the program integration (linkages among the program's sub-components) will appear fragmented, even though the sub-components may appear of high quality individually.

A taxonomy is analogous to a mathematical coordinate system, and the requirements for a high quality S&T taxonomy parallel those of a high quality coordinate system. These requirements/ characteristics are:

IV - 3 - A) Orthogonality - a good coordinate system has orthogonal axes, where the inner product between any two axes is zero. This avoids multiple counting and axis redundancy. Similarly, a good taxonomy should have categories as

independent as possible.

IV - 3 - B) Completeness - a good coordinate system has sufficient degrees of freedom to cover the full range of dimensionality of the physical problem. A 2-D coordinate system would be insufficient for representing a 3-D problem. Similarly, a good program taxonomy will have a sufficient range of categories to include the different technical disciplines that could occur.

IV - 3 - C) Unit basis vectors - a good coordinate system has the unit vector for each dimension the same size. This avoids resolution mis-matches. In addition, the computational grid size should have adequate resolution to allow computational results to be compared to experimental results. Similarly, a good program taxonomy should include technical disciplines of relatively equal importance with relatively equal amounts of funding, with sufficient category resolution to allow equal levels of coherence about a central theme.

IV - 3 - D) Alignment - a good coordinate system is aligned with the structure of the physical problem. This simplifies the solution by reducing the conversion/ translation between the grid and the structure. A spherical coordinate system is more appropriate to representing a spherical body than a cartesian rectangular system. Similarly, a good program taxonomy should be impedance-matched to data availability.

Assume that these guidelines are followed in taxonomy selection for the proposed review, and a taxonomy of forty categories is defined to represent the total program.

IV - 4) Review Panel Selection

The third operational step is review panel selection. The availability of information technology capabilities will allow the following substantial panel enhancements relative to traditional peer review procedures.

IV - 4 - A) Use of Group-Ware for entering data and computing summary rating statistics in real-time will allow a much larger and more representative segment of the technical community to actively participate in the process;

IV - 4 - B) Having a larger panel will allow the expert witness function and the jury function to be de-coupled, similar to the procedure of the Science Court (DOE, 1978);

IV - 4 - C) Having a larger panel will also allow reviewers to be selected with expertise in a particular evaluation criterion for a specific technical area;

IV - 4 - D) Use of data mining techniques in different literatures will allow a larger pool of experts to be identified as potential process participants.

For the proposed review, assume there is a central panel of perhaps fifteen individuals, and there are one hundred expert reviewers. The fifteen central panelists would not necessarily be expert in any of the areas reviewed, but would be high caliber individuals as free as possible of potential conflict with the programs under review. In the legal analogy, they would serve as the jury. The hundred expert reviewers would be divided equally among the five criteria, or twenty per evaluation criterion. In the legal analogy, they would serve as the expert witnesses. While complete independence from the programs reviewed would be preferable for the expert reviewers, it would not be the absolute requirement used for the fifteen central panelists.

The fifteen central panelists would be selected based on national reputation and absence of conflict. Their function would be to provide final ratings and comments on all the evaluation criteria for all forty programs under review. Their inputs would consist of background material provided by the program presenters, actual program presentations, and preliminary comments and ratings by the one hundred expert reviewers.

Expert reviewer selection would proceed as follows, using the Technical Approach/ Payoff criterion as an example. In parallel with recommendations for experts in the forty technical areas under review, the literature would be 'mined' using key phrases that describe the forty technical areas. A large number of reviewer candidates would be obtained. Bibliometrics would be employed to winnow this list through identification of those candidates with extensive publishing and citation records. Other reviewer selection criteria would be

employed, to insure that bright younger people, who have not yet established a publication track record, would be included in the review process. All four of these selection approaches were used to nominate participants for the innovation workshop referred to previously, and have been used in part by the first author for other types of reviews as well.

The twenty candidates selected as expert reviewers for the Technical Approach/ Payoff criterion would have two required output products. They would provide comments and preliminary ratings only on the single evaluation criterion for each of the forty programs. In order not to overwhelm the fifteen central panelists with comments and preliminary ratings from each of the twenty expert reviewers for each of the five criteria for each of the forty programs, one of the expert reviewers for each criterion for each program would be assigned the task of aggregating and summarizing the comments and preliminary ratings for the given criterion and program. To insure a balanced summary is presented from the expert reviewers to the central panelists, another of the expert reviewers for the criterion would have to approve the summary generated by the expert with primary authority. This expert with secondary authority would be selected based on maximum divergence with the viewpoints of the expert with primary authority, to the extent known beforehand. In the illustrative example, each expert reviewer would serve as the primary authority for Technical Approach/ Payoff for two programs, and would serve as the secondary authority for Technical Approach/ Payoff for two other programs.

IV - 5) Operational Review Process

Selection of the goals and objectives, evaluation criteria, review taxonomy, and reviewers, and definition of assignments and responsibilities, establish the structure of the review. The structure, in turn, provides the foundation for the operational review procedure that follows. The complete review process proposed here will consist of three phases: pre-presentation, presentation, post-presentation. The steps emphasized are those in which the use of information technology, especially in the network-centric mode, will enhance the efficiency and quality of the peer review process. Most of the procedures proposed have either been used or tested to some degree by the first author, and their feasibility has been demonstrated.

IV - 5 - A) Pre-Presentation Phase

The objectives of this phase are to provide as much information to all the review participants as is possible before the meeting occurs, and to clarify any outstanding questions and issues. This will allow the participants in the presentation phase to start on a much higher plane, and use the presentation period much more efficiently.

This pre-presentation phase has three distinct sub-phases. First is the distribution of background material. This sub-phase objective is to provide maximal information about the programs to be reviewed and about global efforts in the programs' technical areas and allied disciplines. Since all reviewers are required to provide a preliminary rating on one criterion for every one of the forty programs, this sub-phase will provide the threshold level of understanding about each program necessary for casting an intelligent vote.

The second sub-phase consists of e-mail interaction among reviewers, where comments are exchanged about the program material and issues are clarified. At the end of this sub-phase, each reviewer has transmitted his/ her comments on the assigned evaluation criterion for each of the forty programs to the individuals assigned primary and secondary responsibility for the specific criterion for each program.

The third sub-phase consists of the primary and secondary principals responsible for each criterion for each program writing a brief summary based on the inputs of the other reviewers assigned to each criterion for each program. At the end of this sub-phase, these brief summaries will have been transmitted to the fifteen member central panel, along with the preliminary summary rating statistics for each criterion for each program.

IV - 5 - A - i) Distribution of Background Material

This phase begins with the distribution of background material for the reviewers (and audience, if an audience is desired). In order for the background process to be most effective, the material should be distributed at least three months prior to

the actual presentations. Two types of material are proposed.

First are narratives and vugraphs describing in detail the material to be reviewed. The first author distributes this type of background information routinely for S&T peer reviews. Requirements for this material have been detailed elsewhere (Kostoff, 1998). To maximize distribution efficiency, the material should be made available on the Internet, and the reviewers/ audience informed of its location. If distribution of some of the material has to be restricted for proprietary or other reasons, then the Web site should be password-protected.

The second type of material is information related to the programs to be presented. This material is 'data-mined' from appropriate source S&T databases (e.g., Science Citation Index (basic research), Engineering Compendex (applied research and technology), NTIS Technical Reports (government-sponsored S&T reports), Medline (medical S&T), RADIUS (narratives of on-going government R&D programs). The first author has distributed 'data-mined' information to support reviews of technical areas of modest breadth. This information can be very valuable in identifying the scope of S&T performed globally in the specific technical area under review, in allied areas, and in disparate fields that have some thread of commonality with the specific area under review.

However, even for fields of moderate breadth, substantial effort is required to provide useful background information of this type. The query used has to be refined to satisfy two conditions: the coverage (records retrieved) should be comprehensive (large signal), and have minimal extraneous material (large signal-to-noise). Then, for most recipients, the records retrieved need to be summarized. The first author has used the Database Tomography approach (Kostoff, 1999b) to develop queries with these properties, and to summarize the main pervasive technical themes in such retrieved record databases, and the relationships among these themes. While these computational linguistics and bibliometrics tools help substantially, they do not obviate the need for technical experts to spend substantial time and effort in developing this background material.

For the illustrative example used in this report, a forty sub-program Advanced Technology Development naval S&T program, the effort required for global data

mining of the technical disciplines to be reviewed would be enormous. Nevertheless, if each reviewer's rating is to be meaningful, then the reviewer needs to have some threshold level of understanding about each program reviewed. A substantial effort is necessary to provide such information, especially in summary form.

IV - 5 - A - ii) Individual Reviewer's Comments

The discussion in this sub-section follows the experience of the innovation workshop in Autonomous Flying Systems mentioned previously. Even though the objectives of a workshop are different from those of a peer review, nevertheless, the principles learned from the workshop's pre-presentation phase can be readily extrapolated to peer review application.

In the innovation workshop, each participant sent new concepts relating to the workshop theme to all the other participants by e-mail. An e-mail-based interactive discussion ensued among the participants to 'flesh-out' the concepts, and either clarify and/ or embellish them in preparation for the actual presentations. In order to stimulate this e-mail discussion, a facilitator was required to raise numerous questions. The discussion proved extremely successful in clarifying the concepts, but the need, and effort required, for facilitation of the discussion was appreciated only after the pre-presentation phase had begun.

In this phase of the peer review process, after the reviewers have received the background material, they would be expected to spend the next few weeks digesting the material and clarifying any outstanding or problematic issues. The primary and secondary principals for each criterion for each program would be expected to act as facilitators, to stimulate discussion on these issues. The total review group would not be involved in each e-mail discussion group; this would overwhelm the communication channels. Each e-mail discussion group, in the present example, would consist of the twenty experts for a given evaluation criterion for a given program, plus the individual who will be presenting the information. At the end of this phase, approximately two months before the presentations, each of the twenty experts would provide his/ her comments and preliminary ratings on the given evaluation criterion for the given program to the

appropriate primary and secondary principals.

IV - 5 - A - iii) Summary Comments to Central Panel

After receiving the individual comments and preliminary ratings from each reviewer, the primary and secondary principals for each criterion for each program will generate a brief summary for each criterion for each program. If the two principals cannot agree on a specific summary, the secondary principal will contribute a dissenting addendum to the summary transmitted by the primary principal to the central panel. In any case, both the comment summary and a summary of the preliminary rating statistics are transmitted to each member of the central panel. In order for the central panel members to have time to absorb all the summary material, they would need to receive it no later than one month before the presentations.

In summary, the total pre-presentation time-line is as follows:

*Distribution of background material to expert reviewers - three months before presentations
*Transmission of comments and preliminary ratings to primary and secondary principals - two months before presentations
*Transmission of summary comments and preliminary rating statistics to central panel members - one month before presentations.

IV - 5 - B) Presentation Phase

In network-centric peer review, this phase is optional. There is no fundamental requirement for presentations. All of the review could be conducted through the network by e-mail, Internet, etc. However, there is a cultural aspect to peer review that rivals the information technology aspects in shaping the conduct of the review. Many cultures are not yet at the required comfort level with purely remote operation. In addition, there is value in real-time discourse with the presenters. Therefore, this presentation phase will be included in the present paper.

For the scenario proposed in this paper, presentations will be made to an on-site

audience consisting of the fifteen member central panel and the onehundred member reviewer group. Presentations can also be made to a remote audience by video tele-conferencing. Under the present scenario, the role of the remote audience is observation.

All the members of the on-site audience will be linked by Group-Ware. During the presentations, the reviewers will enter final ratings and any additional comments they believe are important based on last-minute observations or insights. At the end of each presentation day, the remote transmission link will be closed, and the reviewers and central panel will meet in Executive Session. The Group-Ware algorithms will have computed each program's statistics (panel averages for each evaluation criterion rating, etc) and any desired integrative statistics over multiple program groups as well. All these numerical results will be displayed graphically to all the on-site audience. The Group-Ware will have also aggregated the additional comments, and these comments will be displayed to all the participants. Both the ratings and the comments will be discussed for each evaluation criterion for each program presented. The central panel will then rate each evaluation criterion for each program presented, and these final program and integrative statistics will be displayed in real-time.

A note about Group-Ware. In the naval Advanced Technology Development review described in the text, Group-Ware was used in part. It had two components: computing summary and integrative statistics, and aggregating comments. Both these features operated in real-time. The immediate summary and integrative statistics feedback provides for high efficiency discussions, and its value increases as the number of programs reviewed and the number of experts used increase. The comment aggregation is valuable for documentation purposes. For an on-site panel, comment aggregation has little value, can serve to bias reviewers' initial comments, and can be a distraction to some reviewers. For reviewers from remote locations, comment aggregation should prove to be of substantial value.

IV - 5 - C) Post-Presentation Phase

This phase consists of writing the final review report. Depending on the contractual structure of the review, either the staff of the organization sponsoring

the review will write the report, or the central panel will write the report. Because of the extensive pre-presentation preparation, the involvement of a large segment of the community, and the extensive interactions that occurred during all prior phases of the review, much of the available information will be ready for direct insertion into the report.

V - RESEARCH OPPORTUNITIES IN NETWORK-CENTRIC PEER REVIEW

Opportunities for research into network-centric peer review abound. Issues to be addressed include the following:

*How is peer review quality defined, especially in a network-centric mode? What are the metrics of quality; how can they be measured? What data is required to quantify these metrics, and how is this data obtained?

*What incentives and rewards have been employed to produce higher quality reviews, and what incentives and rewards should be tested for efficiency?

*What types of network architectures should be developed for optimal review operation? How extensive should the networks be for successful operation? What are the implications of reviewer anonymity protection on the network architectures? What other types of security and verification procedures are required to minimize review disruption and corruption problems? What levels of fault-tolerance need to be incorporated into the network? What are the hardware and software requirements for optimal large-scale operation?

*What are optimal reviewer selection processes, and what are the trade-offs among these processes?

*What are the cost-benefit considerations related to panel sizes, for different types of review objectives? What are the trade-offs of adding more experts in a given technical area for statistical reliability and validity purposes verses broadening the expertise representation across many different fields? How far should the expertise diverge from the target S&T being evaluated, in order to access insights from other disciplines that could benefit the target discipline?

*What are the trade-offs involved in Science Court operation verses dual function jury-witness panel? What other panel operational modes are possible with network-centric operation? What has been the experience of these other operational modes; what is the potential of other operational modes, whether or not there has been some past history of operation?

*What credible processes exist, or could be devised, to normalize across panels and disciplines? How does network-centric operation complicate or simplify these diverse processes?

*How does the expanded capability of network-centric operation impact the selection of diverse evaluation criteria, and how does it impact the development of, and accession to, the data required to address these criteria?

*How are reliability and repeatability impacted by network-centric operation?

*How should the different types and sources of global data be accessed and integrated with the peer review process? What are the implications on the process operation and results on the availability of these different types of data? What data sources need to be developed and constructed to provide required information for peer reviews, and how does network-centric operation influence the composition and structure of these sources?

*What are the true costs and benefits of network-centric peer review, and what are the main parameters that affect cost-sensitivities? What steps could be instituted now to reduce potential high cost components of the network-centric peer review process?

*How should the larger network-centric strategic management process be constructed in order to maximize benefits from network-centric peer review, as well as optimize benefits organizationally and nationally from the strategic management process? What constraints do the other elements of the network-centric strategic management process place on efficient operation of the network-centric peer review component, and what enhanced capabilities for the peer review component do these other components offer? What are the common elements of all the components of the strategic management process, and what are the unique

elements required for network-centric peer review? Are there benefits to constructing architectures that will encompass all the network-centric strategic management components, such that specific requirements for the peer review component will require a minimal additional requirement for resources?

VI - SUMMARY AND CONCLUSIONS

Network-centric peer review uses the power of modern communication networks and information technology to expand greatly the number of people that can participate in real-time peer reviews, and expands greatly the access to data that can support all aspects of peer review. This technology allows diverse review operational modes such as the Science Court to be considered seriously, and allows the jury function of peer review to be independent from the higher conflict potential expert reviewer/ witness function. The operational architecture required for network-centric peer review may differ little from the architecture required for its parent network-centric strategic management, and since all strategic management components need to be integrated for optimal synergistic benefits, implementation of network-centric peer review should occur in parallel with implementation of the other components of network-centric strategic management.