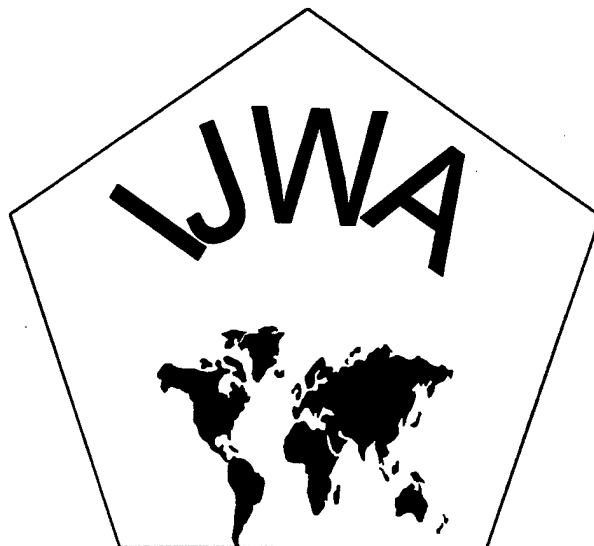# COMPLEX EXPERIMENTATION PROCESSES

FLEET BATTLE EXPERIMENT IMPLEMENTATION

SUMMARY REPORT

Gordon E. Schacher
Shelley P. Gallup, Jr.

January 2001

The Institute for Joint Warfare Analysis
Naval Postgraduate School
Monterey, California

COMPLEX EXPERIMENTATION PROCESSES

FLEET BATTLE EXPERIMENT IMPLEMENTATION

SUMMARY REPORT

Gordon E. Schacher
Shelley P. Gallup, Jr.

Institute for Joint Warfare Analysis
Naval Postgraduate School
Monterey, California

RADM David R. Ellison
Superintendent

Richard Elster
Provost

This report was prepared for and partially funded by:

Institute for Joint Warfare Analysis, Naval Postgraduate School

This report was prepared by:

Institute for Joint Warfare Analysis
Naval Postgraduate School
Monterey, CA

Authors:

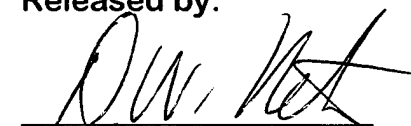_____
Gordon E. Schacher

_____
Shelley P. Gallup, Jr.

Reviewed by:

_____
GORDON E. SCHACHER
Director
Institute for Joint Warfare Analysis

Released by:

_____
DAVID W. NETZER
Associate Provost and
Dean of Research

REPORT DOCUMENTATION PAGE

Form approved
OMB No 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE January 2001	3. REPORT TYPE AND DATES COVERED Technical January 2000 – January 2001

4. TITLE AND SUBTITLE
Complex Experimentation Processes, Fleet Battle Experiment Implementation Summary Report

5. FUNDING

6. AUTHOR(S)
Gordon E. Schacher, Shelley P. Gallup, Jr.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Institute for Joint Warfare Analysis
Naval Postgraduate School
Monterey, CA 93943-5000

8. PERFORMING ORGANIZATION REPORT NUMBER
NPS-IJWA-01-011

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Naval Postgraduate School
Monterey, CA

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
None

11. SUPPLEMENTARY NOTES

The views expressed in this report are those of the author and do not reflect the official policy or position of the Department of Defense of the U.S. Government.

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words.)

This report provides an interim description of methodology and experimentation process emerging in Fleet Battle Experiments. The developing science of complex system experimentation is expanded in this report, and applicable to other DOD related complex systems projects.

14. SUBJECT TERMS
Fleet Battle Experiments, Operations Research, Systems Analysis, Experiment Methods, Complex Systems

15. NUMBER OF PAGES
123

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT UL

Acknowledgements

EXPERIMENTATION UTILIZING OPERATING FORCES FLEET BATTLE EXPERIMENT IMPLEMENTATION

CONTENTS

This page intentionally left blank

*This report is a collection of papers written for those who
Are part of and familiar with the Fleet Battle Experiment
Process. They value is limited for those who are not part
of these experiments because context is not presented.*

I. Introduction

Any type of experimentation, to be successful, requires a great deal of planning for capture of data and subsequent analyses. Both must be linked to a set of learning objectives. There is a progression of types of experiments, from those that are simple to plan to those that tax the most ingenious minds.

At one end is experimentation with physical systems in the laboratory, such as solid-state physics. The experiments can be very complex, requiring accurate instrumentation and a progression of detailed cause-and-effect measurements. However, one is dealing with a limited set of interactions and planning is reasonably straightforward. One comforting fact is that the experiments are repeatable, can be controlled, and statistical validity can be assured.

Engineering systems are more difficult to deal with by virtue of the fact that they are made up of many components, so the interactions are more complex, but one is still dealing with physical systems and the same comments as above apply.

If one moves from the laboratory to the field, experimentation becomes more difficult. This is due to lack of control over the environment. Experimentation in fields such as meteorology and oceanography are inherently more difficult than physics, which is why it has taken much longer to accomplish accurate weather prediction than understand the behavior of electromagnetic waves. Repeatability in field environmental experiments means waiting for the right conditions to occur.

If humans are a part of the experiment, control is very difficult. One has to develop means for accounting for the variability of human behavior, or set up environment controls within which human interactions can be investigated. Often one is attempting to determine the effectiveness of a physical system with which humans interact. Apportioning cause-and-effect between the humans and the system is then the challenge.

Of course, if one is dealing with humans in the field, the most challenging experimental situation is encountered. Military field experiments where one is utilizing operating forces is perhaps the most difficult of all. A root cause is that one wishes to determine the effectiveness for executing military operations but it is not possible to reproduce a true warfare situation. A mixture of live-play and simulation is often a means used to inject as much realism into the experiment as possible. Since utilizing military objects: ships, aircraft, personnel, etc. is a very expensive proposition, many replications will most often not be achieved.

In spite of these difficulties, military field experimentation is successfully carried out. However, success depends critically on process. The purpose of this report is to document the process that has been developed for the Navy's Fleet Battle Experiments (FBEs). The format is a set of

papers dealing with the topical areas that form the full process. They are essentially stand-alone documents which were written by the authors over a six-month period and are presented for the most part as they were written. This means that there is some overlap of concepts between them. A small amount of editing has been done to make the whole report a coherent document, but not much.

The following diagram indicates characteristics of experiments and where FBEs fit within a spectrum of types of experiments. Note that the relative positions of the various characteristics along their vertical axes has no relational meaning.

Relatively   Games   Complex
Un structured Qualitative   Systems with
                                                       Humans

                                    Field
                                    Experiments   Processes
                                                                        FBEs
                                    Lim. Obj. Exp.   Systems of
                                                       Systems
                                    Demonstrations

                                    Test and   Hardware
                                    Evaluation   Systems

Completely   Controlled Lab   Basic
Structured   Quantitative   Experiments   Phenomena
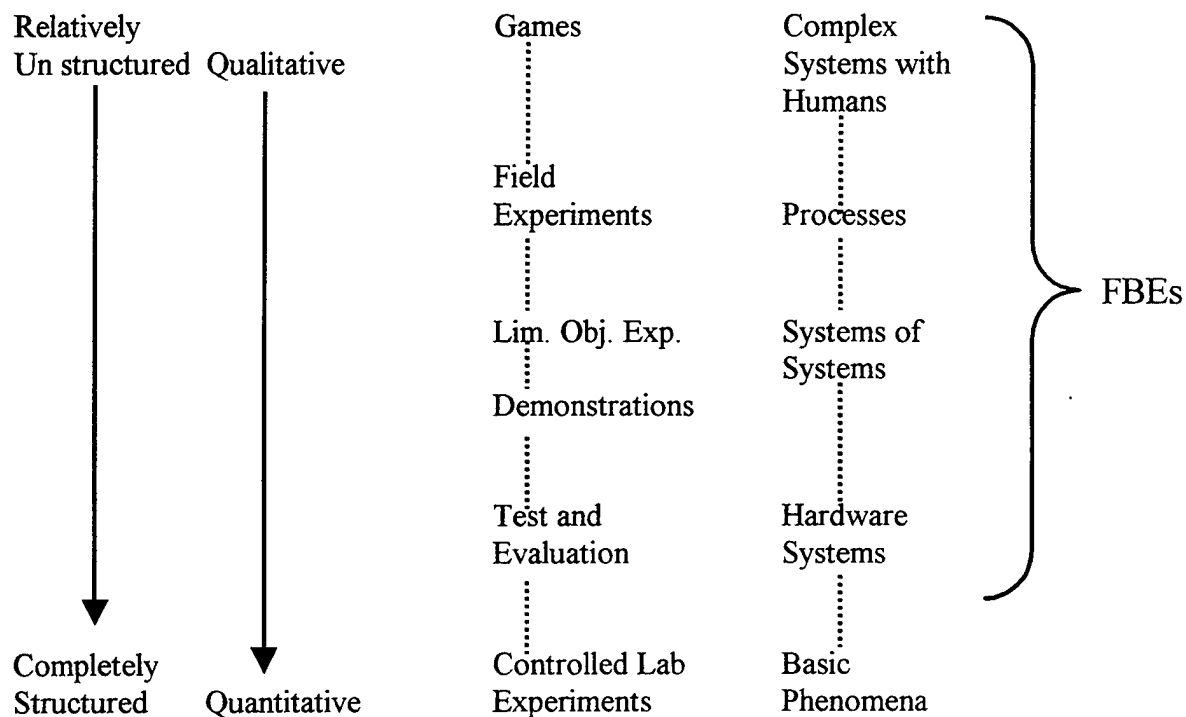
The diagram shows the types of experiments and processes/systems one is dealing with as you move from unstructured, high-complexity experiments that are largely qualitative to those that are highly quantitative with well controlled structure. As the diagram shows, FBEs occupy a large part of the spectrum. This means that a high degree of planning is needed to insure that all of the varied objectives are met.

The next section provides a brief explanation of the purpose of each document, and provides some glue for the whole. One can use these explanations as a guide to moving around through the documents, picking the ones of interest for examination. In general, the ordering of the sections moves from general concepts to specifics.

4

II. Section Descriptions

As was stated above, this report is a series of papers. They have been arranged to progress from general requirements for experimentation to contributions to the transition of military forces to meet new requirements, to specific activities that are needed to carry out field experiments, to reporting of results. Starting at the beginning and reading to the end will present a complete picture, but many readers will wish to jump to specific topics.

This Section contains short descriptions of each of the papers, which should act as a guide for the reader. Note that not all of the papers are independent. In some cases, a topic will be covered in more than one paper and no attempt has been made to combine them into a single section. Thus, some redundancy will be encountered, but not duplication of content, rather expansions or slightly different explanations of a point or process.

Requirements for Organizations Engaged in Technology Products

Organization literature has many case studies that discuss corporations operating in high technology and innovation environments. Successful aircraft companies, computer software companies, universities, etc., all have a similar component structure. This structure is presented and related to structural requirements for Naval Warfare Development Command (NWDC).

Developing Initiatives

FBEs are developed around a number of concepts (usually wide in scope) and initiatives (more targeted to needs). These result from discussion within NWDC and from Fleet concerns. The development of an experiment (apart from the execution of an exercise or other event in which an FBE may be conducted) depends greatly on this umbrella of concepts and the consequent set of an individual experiment's initiatives. The process of developing initiatives is discussed.

Constructing Data Plans

A method is shown for identifying experimental objectives (requirements for knowledge), defining capabilities that fulfill those needs, and elements of data that are specific for data planning and execution. A rigorous data collection plan (DCP) results from this method, and an opportunity to conduct analysis across the continuum of experimentation.

Decomposing Learning Objectives and Experiment Questions

Learning objectives and questions to be addressed in an experiment are most often stated in broad terms. They must be treated as general guidelines from which specifics have to be derived. Decomposition and extracting specific objectives and questions for which experimental data can be obtained are described.

Internal Consistency (Fitness) in Experimentation

Internal consistency between the various aspects of an experiment must be developed if the experiment is to produce results that are meaningful, of high quality, and address the experiment learning goals. A general discussion of the topic and examples are presented.

Process Status Chart
Five parallel processes need to progress in order to construct a Fleet Battle Experiment data-capture plan. The steps in the process are described and a chart is presented for tracking progress.

Concept Centered Experimentation and Analysis
Designing, executing, and analyzing Fleet Battle Experiments (FBE), and ensuring that results can be carried forward to future events, is a complex process. A robust process is needed that includes synergistic games, studies, exercises, etc. building on one another and leading to accepted and well documented results. This paper outlines a Concept Centered Experimentation and Analysis process for FBE formulation, planning, execution, and analysis.

Knowledge Management Structure
Fleet Battle Experiments generate large amounts of various types of information and data. Both must be archived in a way that allows easy extraction to perform various analyses. A multi-level tagging scheme is used to provide access to individual data elements. This allows building sets of tags to pull data that address specific question of study goals. The scheme for subjective data is described.

Data Requirements
Many types of data are needed to capture the full extent of Fleet Battle Experiment operations. One must gather information on processes, systems, human interactions with real-time information, bits and bytes flowing in electronic systems, etc. This paper describes the types of data and how they are gathered.

Reporting Structure
There are many customers for FBE results, some internal to the process, many external. The following notes several of the customers, the types of information they need, and the reporting structure that meets those needs. The reporting structure is supported by a synergistic information structure, which is also briefly described.

Case Study Analysis
Fleet Battle Experiments are not controlled experiments from which one can generate statistically significant results. Rather, they are vignettes from which one can observe a result for a set of circumstances. In essence, they are case studies, and what this implies for how results are reported is presented.

In addition to these papers by the authors, there are germane documents prepared by other authors which appear in appendices. They are:

<u>The Modular Command and Control Evaluation System</u>
The Military Operations Research Society has devised a methodology for evaluating Command and Control systems. An outline of the methodology is presented in this paper. The material is extracted from a summary by the Society

<u>Navy Major-Caliber Ammunition Reliability Goals</u>
This paper describes data capture and analysis for the Navy 5 inch gun program. It is an excellent example of inductive analysis and reporting of results.

<u>Process for Determining Measures of Evaluation</u>
This section is a set of Power Point slides that outline a rigorous methodology for developing measures of performance and effectiveness.
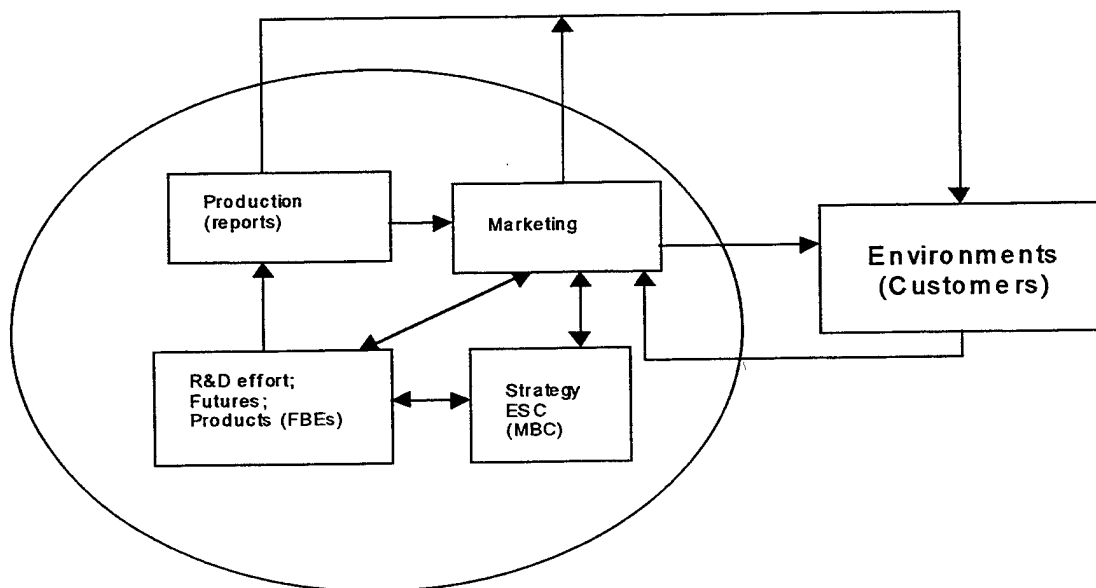
This page intentionally left blank.

III. Requirements for Organizations Engaged in Technology Products

Organizations engaged in high technology and discovery of innovation may be described in some generalized interactions. The components of these interactions are included in the diagram below.

Organizations that are responsible for production of high technology or technical products (such as aircraft companies, computer software companies computer hardware companies and universities, to name just a few) all have a similar structure with at least these components.

Generally there will be (not in any order), a Marketing division, a strategy or oversight organization, a research and development division that also will include some prototyping function, and a production division. Of course there are other components to these complex

```
                                    ┌─────────────────────┐
        ┌───────────────┐           │                     │
        │  Production    │          │   Marketing         │
        │  (reports)     │────────▶ │                     │────────▶  ┌──────────────────┐
        └───────────────┘           └─────────────────────┘           │  Environments    │
                                                                       │  (Customers)     │
        ┌───────────────┐           ┌─────────────┐                    └──────────────────┘
        │  R&D effort;  │           │  Strategy   │
        │  Futures;     │◀────────▶ │  ESC        │
        │  Products (FBEs)│         │  (MBC)      │
        └───────────────┘           └─────────────┘
```

organizations that are not included here (such as a financial department, personnel, facilities etc) because they are less value to this discussion.

The oval above marks the boundary of the "system" from the point of view of the organization. Somewhere beyond the boundary of the company is an "environment," which is actually a complex set of environments. The tendency is almost always to think of customers, competitors, problems and so forth as outside of the organization's boundaries, and as part of "the" environment. The result is there is often a great deal of ambiguity as to what the "environment" is to any one person within the organization, unless the organization has spent a great deal of time and effort to focus meanings of "environment." This has its own set of problems, because the collection within "environments" is very dynamic. For the rest of this discussion, "environment" is referred to in the collective sense.

Within the boundaries of the organization (inside the oval), there is very often tension between "marketing" and "production." One major reason for this results from marketing's incentives. One of marketing's roles is to assess and understand what impact the organization has on its targeted "environment," and vice-versa. If marketing is responsible for directing the organization's impact on the environment, it would imply that production and R&D is somehow responsive to marketing. However, it is generally true that R&D will lag behind marketing needs because marketing does not strategize the future of the organization. Strategizing is the role of the organization's leadership, and although that leadership will acknowledge marketing's needs, marketing will not be the central feature of the organization. Strategy's job is difficult because of the resources that have to be continually coupled to the environment. Further system lag is created by the time it takes to produce prototypes or other products. The impact is felt on production, which is not directly responsible for R&D, marketing and strategy, but deeply coupled to all of them.

Most organizations allow for a certain level of this tension, and try to use it to the organization's best advantage.

One outcome of this tension is that marketing, unable to get R&D and production to shift in order to meet changes in the environment (remember we are talking about a high tech company here--not making hamburgers—and even that endeavor has difficulties keeping up with market changes), will attempt to "spin" production's efforts (the product) to make it appear to meet environment's requirements.

Of course there are many permutations to this set of dynamics and interactions.

Experimentation as a Technology Organization

There is a great deal of ambiguity in the environment. Lack of definition here contributes to conflict in the organization as a whole. This is evidenced by a pendulum swing between concerns for "innovation" to "impacting the acquisition process," which has surfaced alongside of former discussions such as "value added to the operations personnel," and "creating buy-in within the flag community." All of these are meaningful to the organization, and are perhaps not without basis, but are also shifts to match perceptions of change in the larger environment described in the model above. The result is that the experimentation organization has tended to focus on one of these customer groups at a time, and at different times. Within each of these potential environments of interest, there is very little common agreement as to what exactly defines each. Without common definitions, it is difficult to combine efforts that result in desired impacts.

A result of the swinging pendulum of priorities mentioned above is that there are impacts on core principles. While these may change, they do not change quickly. So, while the environment seems to demand one category of experimentation results (for example acquisition related data), the organization may have focused on innovation and will experience internal fracturing of the means to produce a result for ambiguous ends. Carrying this example forward, while innovations research is future oriented, marketing (responding to perceived need for results that

are useful to test and evaluations related to acquisitions programs) will not be very interested in designing for futures research and the focus will be much more on present planning.

Marketing is not yet a core competency of experimentation organizations. There is very little formal assessment of environments, and little formal discussion above the anecdotal level of marketing needs. In other words, there tends to be a great deal of ambiguity about who the customers of experimentation are.

Ambiguity and lack of marketing competency will produce a reliance on R&D and production to fill this void. For example, there could be pressure on R&D to create immediate relevance in what was once "innovations research." There will be multiple impacts throughout the organization dynamics as a result.

Recommendations

A "marketing" component needs to be included in experimentation that functions to articulate the *multiple* needs of complex environments and which can also evaluate changes in the environment and assist in integrating change within the rest of the organization.

R&D must have a range of autonomy consistent with functions of research and development. Prototypes always have one foot in grounded requirements and one foot in innovation (looking towards the next prototype). A good R&D function will also produce multiple prototypes that may or may not be directly mapped to marketing needs to impact the environment.

This page intentionally left blank.

IV. Developing Initiatives

BACKGROUND

FBEs are developed around a number of concepts (usually wide in scope) and initiatives (more targeted to needs). These result from discussion within NWDC (internal dialogue) and external Flag (Fleet) concerns. Concepts and initiatives have been problematic in the FBE planning and experiment definition processes, as they are generally not fully formed at the beginning of an experiment planning process. Instead, concepts and initiatives are fleshed and solidified in parallel with the planning process. Yet, the development of an experiment (apart from the execution of an exercise or other event in which an FBE may be conducted) depends greatly on this umbrella of concepts and the consequent set of an individual experiments' initiatives.

Preceding FBEs that were not well defined in concept and supporting initiatives experienced excessive scope creep. Boundary definition is critical to planning, execution and data collection, and to final analysis. Freezing boundaries is especially important as the experiment draws closer to execution. If not properly defined, initiative efforts (the leads and technologies) will continue to define architectures up to execution. This creates ambiguous conditions for everything that follows.

CONCEPTS-TO-INITIATIVE DEVELOPMENT

Initiatives generally begin as concepts at NWDC, or from Fleet concerns. Concepts are often vague, encompassing a much broader scope than can be engaged in one experiment. The model to date has been to loosely define a set of concepts (combining those of both NWDC and the Fleet) on which there may or may not be consensus between NWDC and the Fleet. A Concept Development Conference (CDC) has been one means of negotiating experiment boundaries by defining "edges" to concepts and supporting initiatives, as well as a means to bring a third component (additional stakeholders) to the mix. Rather than providing structure to the process, the CDC has generally produced increased experiment (still a "project" at this time, and not an experiment) scope. This boundary may or may not have been reconsidered by the time the fourth component of the experiment process, technology, is added in a "technology conference."

There has been little or no opportunity to explicitly state what is proposed to be learned in this process. This cannot happen until the set of initiatives, technologies and experiment methodology are brought together to co-evolve the experiment. Instead, parallel processing with limited scope review has occurred in past experiments. This resulted in delays to "freezing" each of the components (concept, initiative, technology) essential to final description of experiment architectures and analytic methodology.

ITRATIVE PROCESS TO DEFINE INITIATIVES

A lineage should exist between high level NWDC/Fleet concerns (concepts and critical issues), and sub-initiatives, questions that are being answered or knowledge added to, experiment methodology, data definition, analysis and finally—learning. This is the experiment definition framework.

The first phase, developing the initiatives (critical issues) from wider concepts, requires that some iterative dialogue take place between NWDC and principle stakeholders. This dialogue may be improved by providing some structure to the experiment definition framework, which may then be used as the central point of discussion between experiment planners (NWDC) and fleet stakeholders.

RECOMMENDATION

For each concept, decompose to a principle initiative, sub initiatives and sets of impact statements:

Concept: (define in general terms in a page or less) "Concept Papers" have been employed in previous experiment planning. However, these efforts have often been diluted by competing definitions of concept paper's roles. It would be best to loosely describe what is meant in a concept, allowing focus to be developed in initiative and sub-initiative explanations. Concepts should be easily negotiated because their boundaries are necessarily ambiguous. Initiatives and sub-initiatives will require more definition, and therefore greater efforts to reach consensus.

Example: At the concept level, "A common operational picture is the principle and most important means by which the CJTF gains situational awareness within the operational and tactical activities taking place within the battlespace, providing similar SA throughout the chain of command. As such the COP is most important to providing dynamic information that is scaleable and may be focused on specific needs within the battlespace. The COP is one, very important element of a FNC. It must have the properties of accuracy, timeliness and usefulness to each user."

Initiative: Provide a COP to the CJTF, Battle watch captains and each echelon of command that is capable of being focused on the specific needs of the commander in that echelon, dynamic to the battlespace, and which provides the SA necessary to deal with the range of requirements encountered at that level of command.

Sub-initiative (1): Use of COP synchronization tools to provide requisite timeliness to the COP. The experimental question here arises from the complexity of information feeding the COP, which must be refreshed within the lowest dwell times of TCT, while also providing access to information about the battlespace at large. In addition, there is some ambiguity with regard to the ability of the tool to impact the COPs refreshing of incorrect or untimely information.

Iteration of these concept-to-initiative descriptions means that they are shared between NWDC and the fleet stakeholder on a routine basis. If a CDC meeting is held, that meeting should be grounded in a set of fairly well defined concepts and initiatives. In other words, the CDC should be an information venue, not a negotiation of concepts and initiatives by a wider audience of competing interests.

IMPACTS ON FURTHER EXPERIMENT DEVELOPMENT

As noted above, defining experiment boundaries is a critical requirement. This cannot occur without first defining concepts and initiatives.

Supporting architectures should not be considered in these discussions. Concepts and initiatives should describe what is *required*, i.e., a set of needs. Architecture and experiment design follow, but need some definition rigor.

To date, experiment definition below the level of concept and initiative has been continuously iterative, with little feedback to stakeholders or other experiment developers (technology providers, installers, M&S, data collection and analysis) unless they have been *inside* of the process directly.

From concept and initiative development to experiment definition is a very wide gap. A first step across this gap is use of formal systems analysis tools to produce *context, entity-relationship, and process* views of what is being proposed. Employing this rigor will assist in creating "fitness" between concept/initiatives, architectures and experiment design.

This page intentionally left blank.

V. Constructing Data Plans

Methodological fitness requires that experimental objectives (requirements for knowledge) imply capabilities to fulfill those needs, and related data elements. These data elements then specify plans to acquire data and execution of that plan. A rigorous data collection plan (DCP) results from data fitness and provides an opportunity to conduct analysis across the continuum (see the definition of experimentation continuum in this document) of experimentation.

The means to construct fitness in complex experimentation follows, and is presented as a series of steps.

PROCESS STEPS

1. Construct a data matrix.

First, experiment initiatives must be specified from the range of initiatives possible. (I)

-----→ these initiatives imply a set of sub-initiatives ($I_{i1,i2,i3}$...)

---→for each sub-initiative there are general questions (Q) possible, which define concept scope.

Also, further refining this set of questions yields:

---→*experimentable* questions ($q_{1,2,3,}$) which represent focused experiment scope.

These questions point to:

---→data that satisfies requirements, e.g., answers a question.

2. There is a continuum of activities in an FBE, which together create an "experiment." This continuum generally includes elements of :

- Demonstrations
- Evaluation
- Innovation and exploration

Each of which may be coupled to one or a combination of:

- Process
- Technology
- Function
- Capability

And each of these may be satisfied by one or more of a:

- System data element (S)
- Technology data element (T)
- Process relationships (P)
- Evaluation MOE (M)
- Unintended data (U)

3. From the above, a matrix is constructed for each q_i. Cells are labeled according to required data element needs (again, these are related to the question being answered).

	Process	Technology	Function	Capability
Demonstration	S/P		T	.
Evaluation				MOE
Innovation or exploration	P			

4. Any combination of data requirements is possible, as defined in experiment scope and questions being asked. Next, it is necessary to specify the meanings for the various data element needs in each matrix. This is done by answering questions such as:

Q1: Is the process defined? Is the architecture, C2 or technology interface complete?
Q2: What experimental elements are available from technology included in the experiment?
Q3: What evidence is there that a function is being performed?
Q4: If the objective is an evaluation, what is the measure of performance? What evidence satisfies comparative intent of the measure of performance?

5. It is now possible to write a question/data element relationship. For each initiative (I):

I-----→Q----→q----→ Knowledge Data (KD) = S and P (cell 1,1 above)--→ specific data (defined in questions above).

EXAMPLE

Initiative (I): Navy Fires

Sub-Initiative (I_1): Joint Battle-space Management
   Definition: (multiple meanings) I_{11} Visibility to all participants throughout the battle-space with regard to movement, employment and availability of assets. I_{12} Dynamic responsibility for assets on targets, and deconfliction. I_{13} Employment of TACAIR via an E2C with LAWS functions and TIBS, coordinated with Joint Fires architecture and processes (4 dimension deconfliction).

Q1. How is "visibility" integrated into the battlespace, and will it have a positive impact on the domain?

   q1. What are data elements of "visibility?"

technical: all data elements are held within and available to decision systems. Refresh rates are acceptable to creating decisions within dwell times. Information is accurate. Data required:

- tracked data at specific nodes (event data resulting from a distinction being made that something in the environment is necessary to be part of a common data stream—may be a COP or may not). Data stream is included throughout the Fires Network. Nodes at LAWS, GISRS, JCSE, TES.
- Data streams accessible at all levels of the decision and weapons targeting command.

process: systems are able to use information. System components are identifiable. Decision rules are useful and incorporate system information.
Process data required include:

- Process to make distinctions in sensed environment that are/are not of interest.
- Process to share information between systems.

q2. Data elements related to positive contribution.

technical: battle-space information is interoperable with systems and processes that use the data. Identification of nodes at which this contribution would be noted.

process: contribution/constraint and relationship to other processes.

Q2. How is "visibility" accomplished in general to all levels of the battlespace?

Q3. I_1 implies a COP function. I_{12} implies a requirement for a COP function and I_{13} is a COP technical capability. In what process are these related?

Just for example sake, the following matrix can be constructed with regard to Q1, q1:

	Process	Technology	Function	Capability
Demonstration	S/P		T	
Evaluation				MOE
Innovation	P			

This page intentionally left blank.

VI. Decomposing Learning Objectives and Experiment Questions

LEARNING OBJECTIVES

Learning objectives are the highest level consideration for an experiment. Consider two types of objectives:

Time Sensitive Targeting – determine the ability to accomplish this task.

C3I System – assess the systems performance.

The first refers to a specific mission, the second to a support system. For each, a great deal of context is needed in order to understand how to implement the learning objective. For TST one needs to know such things as:

weapon and sensor mix,
opposition force lay-down,
scenarios to test,
operational objectives, etc.

For C3I one needs to know such things as:

command and decision making structure,
system architecture,
whether human decision making is considered,
decision/information threads being examined, etc.

TST is a process that can be decomposed, and the learning objective has to be decomposed. "Ability" may refer to the

rate of prosecution of targets,
number of targets that can be simultaneously engaged,
or both, plus other measures.

Since TST is a process, it is decomposed into sub-processes, such as:

Detect > Info Fusion > **Recognize** > Mensurate >
> weapon/target pair > Decide > **Engage** > BDA

Two sub-processes are in bold because they involve human decision-making as a fundamental part of the process and obtaining data about the process involves knowing the human state as well as bits and bytes type information.

Obtaining information about a process normally requires gathering sub-process information in order to develop cause-and-effect relationships. It is seldom sufficient to determine only information about the total process. Thus, process decomposition and determining which sub-process data to gather to meet specific learning objectives is a central part of the planning process.

Considering some of the above sub-processes, one can readily envision types of information that will be desired:

Detect – what is the detection efficiency?

Mensuration – what quality images are needed in reference libraries?

Weapon/Target Pairing – what fraction of targets are not engaged because an appropriate TLE cannot be produced?

Note that the results for the first and last questions can be expressed as numerical measures of performance.

Note that each of theses areas of desired information were expressed as questions. It is often efficient to express information needs as questions. Question decomposition follows.

EXPERIMENT QUESTIONS

High-Level Questions begin the process of focussing the experimentation areas. The questions also illuminate the interests of the experiment sponsors, NWDC and the Fleet Commander. Moving to the next level of questions requires parallel development of information about architectures and scenarios.

Each high-level question will normally be broken down into several sub-questions which are amenable to experimentation. Consider a series of progressively more explicit questions.
1. Can we accomplish our Time Critical Targeting mission?
2. Does a centralized ISR desk in an FBE Cell improve our ability to prosecute TCTs?
3. Can we engage Time Critical Targets within 10 min of first detection?
4. What are the PTW+ mensuration times as a function of target and environment.

A data capture plan **cannot** be put in place for question 1. The question is broad and has the semantic difficulty that "accomplish" is not defined. However, the question can be posed to experts and their opinions can provide some information.

Question 2 has a great deal of complexity. The word "improve" implies that comparison is to be made to a former state, the answer to which would require that baseline information exist or be developed. Ability is not a defined word in an experimental context. A measure is needed, such as a prosecution time, or perhaps the ability to discriminate targets from the background with a new process. The question is not amenable to experimentation as stated. Part of the experiment planning process is to take questions such as this and massage them into one or more **experimentatable questions**.

Question 3 is a **"good"** question. It asks for a specific time parameter and defines the ends of the process across which time is to be measured. Data can be captured for this question. Actually, one might wish to put in place several measurements in order to capture the processing times for various portions of the system. Even though the question didn't ask for a sub-system breakdown, setting up the data capture to allow it to be done would undoubtedly be useful for future detailed analysis.

Question 4 is a sub-component of question 3 in that it refers to a specific part of the system.

There is one further consideration. One should specify the configuration of the tactical system when the measurements are made. It is often the case that configurations change during an experiment, so a measurement has little meaning without a specification of "context", the system state.

SPECIFIC QUESTION DECOMPOSITION EXAMPLE

Consider the following question with regard to time critical targeting:

> "Determine a sufficient level of ISR information and automated
> fusion capability that would permit engagement nodes to
> successfully acquire, identify, mensurate, engage and provide
> initial Battle Damage Assessment (BDA) on a target in support
> of maritime and land attack operations."

This question has many elements. In order to use it for experimentation and analysis design, it has to be decomposed. We have broken it down into the following elements:

Basic Questions – needed information level
                         needed automated fusion capability

Context – Land Attack Operations
              Maritime Operations

Data Nodes – Automated Fusion
                 Target Acquisition
                 Engagement
                 Mensuration
                 Identification
                 Battle Damage Assessment

Information is needed during Maritime support for Land Attack Operations. The basic goals are to determine needed quantities, which is not a well-bounded learning objective since the requirement will depend on the magnitude of the threat. A better-posed objective would be to determine whether the capabilities being tested are sufficient (rather than what is needed which can be difficult to determine) for a particular threat level. The reason for the difference in difficulty in determining needs and sufficiency is that one probably cannot determine needs if the system being used for the experiment is insufficient to the task.

The question addresses several processes from which information is needed (he data nodes). The implication is that sufficiency is needed for all of these processes if the total system is to meet the requirement (a reasonable assumption). One has to determine both whether information can be collected from those nodes and how.

For this example, we look at the requirements for Automated Fusion data. Information is needed about each of the processes before data capture can be designed. Consider automated fusion. One needs information about:

> System architecture
> Sensors data characteristics
> Fusion rules and Processes
> C2 Processes
> Sensor Control
> Etc.

We go further into the decomposition by considering the system architecture. Of course, the architecture has to be well defined before specific data-capture plans can be formulated. However, even with out that information, one can now link specific types of data to gather with the original questions, such as

> Is the bandwidth sufficient to transmit the required imagery information?

This question must also be placed in context, now at a much more detailed level. Context information needed are:

> Number of targets per unit time
> Types of sensors (data streams generated)

Each of the learning objectives can be decomposed in the same way. Fortunately, many of the decompositions will lead to some of the same data requirements, such as bandwidth.

VII. Internal Consistency (Fitness) in Experimentation

Philosophy, the first science, defines principles as "irreducible" and internally consistent statements which have no requirement for further reduction and may be generalized over any subject area. If it is a principle, it doesn't matter where the statement is applied—it will fit.

Internal consistency is a "principle" of experimentation. That is, it does not matter what kind of experimentation one is talking about, whether it is looking for new sub atomic particles, trying to understand an ecosystem, or learning from complex military systems. "Fitness," the essence of internal consistency, is an absolute. That is, without it, experimentation fails to provide the logical relations to make it "science."

The canon of the philosophy of science is huge with regard to this topic, although the terminology varies from author to author. Still, the principle stands, and it is simply this:

> In order for an experiment to produce credible results there must be
> a chain of logic which connects
> > what is to be learned,
> > the means to learn it, and
> > the environment in which the experiment is conducted,
> in such a way that biases may be surfaced so that they are observable
> and may be accounted for.

Two schools emerged in western philosophy of science, both of which claimed to provide the conditions for experimentation under this basic principle. Empiricist philosophy holds that an objective experiment is possible, meaning that it is possible to design experiments in which the observer is not part of the bias. Much of western science was built on this principle, but which has also experienced a great deal of difficulty in quantum physics (where one would have expected the highest degree of empirical control). Here the observer was found to have a great deal of impact on the observations, a very confusing result for science of the 20th century.

Experiential philosophy says the opposite: the observer and the observed are part of the same system. Understanding that system perspective has led to cybernetics and complex systems science.

The reason for the above discussion is to make clear that there are distinctions in science about how science is to be performed. In the empiricist case, systems are not looked at as systems, but are purposely pulled apart to consider the smallest elements that would make up the system. This view has often worked well with experimentation at the phenomena level. Experiential science asks one to consider that it is relations between things, and not the things themselves that are important. This seems to work best at the complex systems level.

The direct impact on complex experiment design is that regardless of one's science philosophy, ensuring **fitness** between
> what is to be learned in an experiment,
> methodology to create the conditions for learning,

means for observing the distinctions that are the data the
experiment is designed to surface, and
a set of logics that create "knowledge" that is "truthful" and related
to the fundamental notions the experiment was designed to explore.
Whether empirical, or experiential, the principle of internal coherence (fitness) has been applied in order to conduct meaningful science.

Application to complex experimentation

Military experiments (e.g., Navy Fleet Battle Experiments--FBEs) are "complex" experiments. This label reflects a common (trivial) understanding of activities as being "complicated." However, in experimentation complicated has a significant meaning, which is important. Complexity in experimentation refers to the notion of a continuum of experiment needs.

Answering the question "what is expected to be learned from doing X," is a necessary precondition for all that follows. For example:
   a) Testing and evaluating a piece of equipment or a technology is an experiment
   category (which is fundamentally different from phenomenological research).
   b) technology interaction testing,
   c) systems experimentation (a system here includes people, technology,
   processes, data flow and rule sets), and
   d) learning from systems of systems interactions.

All of these categories are included in Fleet Battle Experiments. And, to make the final point here, *these categories are interrelated in Fleet Battle Experiments.* This means that standard empirical practices, while appropriate at technology levels of FBE experimentation, are not appropriate at the complex end of the experiment category continuum. In addition, experiential methods are not likely to be the means of acquiring knowledge at the technology T&E end of the continuum (except as human factors are included).

Designing Experimentation Fitness

It is necessary to develop experiments as a dialogue between the multiple needs of the categories included in the experiment. Experimentation fitness is the objective of this dialogue and should follow its own methodology.

Establishing fitness is necessary both within an individual category of experiment, and across categories. Also, it is important to keep in mind that the experiment design process itself is also a "system." This poses a new set of challenges to the experiment design team. A very well developed and grounded methodology, which has been developed for systems work, is the Systems Development Life Cycle (SDLC) approach. A systems methodology to approach experiment design, includes project description, scope development, survey, analysis, design, implementation and support phases.

Experimentation fitness requires that these phases (and milestones within them) are specifically articulated within the "experiment design system (planning)," "knowledge definition and

acquisition system (data collection and analysis)," and at each level of the experiment continuum for which there is an articulated objective.

In general this means that there must be a fit between what is to be learned and:
- Concepts (experiment theme or themes)
- Systems of systems (complex systems) necessary to providing the conditions to "experience" the concept
- Individual systems within complex systems
- Interrelations between systems
- Technology experiments and demonstrations
- Scenarios/events that are "played"

Creating fitness therefore requires that the systems approach (SDLC) must include these considerations. Again, a "system" here refers to people, technology, processes, data flow and rule sets.

The experimentation design system and the knowledge system must include the means to attain experiment fitness. These are process steps which include:

- Defining concepts (themes) and freezing them.
- Developing experiment scope
- Defining what may be learned within this scope (scope must be constantly revisited to ensure this)
- Deciding what the significant questions are, and their fit to themes and scope—these must be frozen in the process.
- Explicitly stating the experimentable questions that are consistent with what is to be learned and experiment scope.
- Describing the explicit data set (this must be precise)
- Plan for means to acquire explicit data, and ensure means to conduct reduction and analysis.
- Compare data (post execution) with themes, objectives and explicit questions for coherence.
- Data reduction and analysis.
- Reporting and feedback to further system and experimentation design.

Implications

Failure to develop the correct fit in experimentation produces loose coupling between the experiment elements in the continuum of experiment categories, ambiguity in experiment scope, limited and ambiguous learning and low quality of feedback for development of further experimentation. Using a systems design approach across the continuum of experimentation types will assist in including notions of fitness within experimentation design, knowledge design and experiment execution. Coherence in complex experiments does not simply "emerge," but is a design criteria.

EXAMPLES

The above lays down the principles of internal consistency in experimentation. The description was general, and necessarily somewhat esoteric. The following are simple examples to illustrate the principles. These examples only treat narrow aspects of fitness. Fitness across the full spectrum of an experiment is much more complicated than these examples illustrate. The desired illustrations are made by presenting examples of non-fitness.

Fixing Experiment Conditions

This is not an example but a general principle that has been alluded to above. It is not possible to insure fitness if experiment conditions are not fixed. There is a tendency in operational experiments to add systems or processes right up to shortly before an experiment begins. It may feel satisfying to try the latest gadget at the last moment, or to satisfy the emerging idea of a person of influence, but doing so prevents insuring a quality experiment.

Time Critical Strike Capabilities

Assume a learning objective to determine if a new process allows rapid prosecution of TCTs. It may be that it is sufficient to determine if the process can work rapidly for a single target. Or, it may be that the goal is to determine the capability in a target rich, stressful environment. If the scenario is target poor, the first objective can be met but the second cannot. The goal and the scenario must be compatible.

Flattened C2 Structure

A goal of the experiment may be to test a distributed C2 system with the far-flung nodes given decision-making authority. If the Task-Force Commander is not comfortable allowing subordinate commanders to do independent decision making, and imposes restrictions such as command-by-negation, and imposes it frequently, no information may be obtained about the capabilities of the distributed C2 process. It may be that the problem with the experiment design was that the Commander was not provided with adequate situation awareness tools so that he could develop confidence in subordinates decisions, or it may be that he was not in agreement with this goal of the experiment.

Effects Based Operations

A goal could be to determine if an effects-based operation, with an effects coordination cell could save resources. If no near-real-time feedback is provided so that tactical effects can be monitored, it is not possible to suspend an operation and reallocate resources to achieve the hoped for efficiencies. This could be a problem with the information system or it could be lack of process planning so that all parts of the planning/execution cycle could adjust rapidly to changing circumstances.

Stimulated Play vs Live Play

When play is stimulated by a simulation, special circumstances arise. The simulation runs on a model, which contains a set of physical assumptions about the behavior of systems. These physical assumptions must also be compatible with experiment goals. For a simplistic example, it is not possible to determine the Pk of a weapon if that is a parameter of the model (it is pre-determined). On the other hand, if the overall Pk of a process is desired, the simulation can be an effective tool for providing targets and weapons effects for the process.

VIII. Process Status Chart

Planning the data capture and analysis for a Fleet Battle Experiment is accomplished by five parallel, interdependent processes. Attached is a status chart that tracks the progress of those processes. The standard green, yellow, and red are used to indicate status. The following text provides an explanation of the processes and the various status blocks. *The basic assumption here is that a completion of the tasks within each of these processes is required before data-capture and analysis plans can be put in place.*

This chart is specifically designed for the data capture and analysis process. It is not meant to represent the status of any other part of planning the exercise.

A chart is required for each of the initiatives in an experiment. The particular chart attached here is colored to show the current status of the Time Critical Targeting initiative.

Process 1 is the control function. The objectives are established and detailed data and analysis plans put in place. The Initiative Lead controls the experiment planning though this process.

Process 2 is the design of the operational and tactical play during the experiment.

Process 3 is concerned with hardware systems, especially those that support C2 information management, dissemination, and display.

Process 4 is the full set of C3I processes. It includes the decision processes and nodes, TTPs, and human interactions.

Process 5 deals with the simulation stimulated part of the experiment play. It is concerned both with the stimulation of events and with the physical modeling within the simulation.

Red Cell deals with the Red objects that will be played during the experiment: types, numbers, tactics, and locations. It is a process in its own right, but for diagrammatic convenience we show it as a supporting process for Play and Simulation.

EXPLANATION OF STATUS BLOCKS

The Chart blocks are colored when reporting experiment status. The meanings of the colors are:
    Green – sufficient information is available to develop detailed plans.
    Yellow – information incomplete, some preliminary planing can be done
    Red – insufficient information for planning
Note that planning proceeds regardless of the status of these blocks, but cannot be completed until the processes are complete.
    Decision Points
Each diamond is a decision point dealing with a particular process and level of planning. If the decision is positive it means that the level of planning is complete and the content is acceptable *with respect to development of the data and analysis plans.*

"Lead OK" means the Initiative Lead makes the decision.
"Anal OK" means that the analysis lead makes the decision.
"Satisfy" is a collective decision.
In all cases the decisions are made in collaboration with the concerned people in the process. The reason for Lead being designated in Process 1 is that a formal control mechanism is needed to insure the Lead's goals are being addressed. The reason for Analysis being designated as primary for Processes 3, 4, and 5 is that electronic data, simulation content, and C3I are tightly coupled to data capture planning.

Areas of Interest are the general areas that will be addressed within an initiative, such as experimenting with the concept of an ISR desk within a Fires Support Element for TCT.

General Plans are the type of events that will be played.

Systems is the full set of hardware, including pipelines, that will be utilized.

Experiment Learning Objectives are the broad areas within which information is desired, such as methods for accomplishing BDA or deconfliction, or timeline reduction.

Note that the learning objectives require information about the systems and general play that are envisioned (dark arrows with octagonal labels).

Scenarios are more detailed descriptions of the play, such as day 2 will be devoted to SEAD and SAM sites will be attacked. It also contains such information as the number of sites and perhaps that pop-up TELs will be used included as TCTs.

Development of the Scenarios requires information from the Red Cell, as shown.

Note that the decision as to whether the Scenarios are satisfactory requires information from the Experiment Learning Objectives.

MSELs are the detailed scripts for the play. They require more detailed Red Cell input.

Note that it is logical to have iteration with other processes for MSEL development. This will undoubtedly occur, but it is not shown on this diagram because it is not a requirement (albeit advisable) for development of the data and analysis plans. The same is true for many other iterations that could be shown in the diagram.

Architecture is what systems will be in place and how they will be wired into a total system. No feedback is shown for architecture design in this chart because data and analysis design play little or no role.

Simulation is used to stimulate non-live play and also to determine the effects of some actions. The important information about the simulation are the events that will be provided, how effects will be played, and the physical assumptions behind the underlying model. Red Cell information is needed to determine what objects and play sequences are needed in the simulation.

Stimulation by the simulation is a core part of the experiment. The decision about whether it will meet experiment objectives requires knowledge of the detailed questions being asked, shown in the chart as the dark arrow from Process 1.

Note that the decision about whether Simulation meets requirements (Anal OK diamond) involves both the questions being asked and the simulation details. *A crucial point here is that this is not a judgement of the simulation, but about whether the simulation and the detailed questions are mutually supportive.* This is illustrated here by feedback from this decision point to both Simulation and Detailed Questions.

Detailed Questions are designed to focus on the details of what information is to be extracted from the experiment. These questions, or statements of detailed objectives, determine which of the Data Elements, MOPs will be captured.

Note in the chart the interplay between the questions, the simulation, and the C3I processes.

C3I Process is the collection of command elements, C2 structure, decision processes, supporting information, perhaps stated in TTPs or CONOPS, perhaps not. It is highly interactive with the supporting architecture.

There is a requirement that the questions being asked and the C3I Process be mutually supportive. Again, this is shown as feedback in the chart.

Data Elements, MOPs is an extensive list of data that could be collected to support the goals of an initiative. The list includes measures that could be used to assess system performance.

Electronic Data Capture is an essential element of an experiment. Its main function is to establish accurate timelines. The capture plan is a portion of the overall data capture plan.

Data and Analysis Plan is the detailed plan that includes what data, by what capture means, at which experiment nodes, and what information will be produced from analysis.

Note that development of these plans requires information from all of the other processes. A key point is that the processes have to be frozen before the final plan can be produced.

Develop MOEs At the same time as the detailed data and analysis plans are being developed, preliminary decisions can be made about what MOEs will be extracted from the data.

Uncover MOEs During the course of the analysis, natural means for presenting results will emerge. This will expend the set of MOEs that is used.

Following is the Status Chart for the FBE data and analysis processes.

IX. Concept Centered Experimentation and Analysis

Designing, executing, and analyzing Fleet Battle Experiments (FBEs), and ensuring that results can be carried forward to future events, is a complex process. A robust process is needed that includes synergistic games, studies, exercises, etc. building on one another and leading to accepted and well documented results. This paper outlines a Concept Centered Experimentation and Analysis process for FBE formulation, planning, execution, and analysis.

The processes described here are being put in place for FBEs but not for associated events. Learning through real-time innovation is an important part of these experiments and application of these processes will not restrict innovation, rather it is a principle of this paper that the inductive process developed through past FBEs be extended and nurtured in future experiments.

Following these processes will result in the myriad processes and organizations involved in FBE planning, execution, and analysis being synchronized and that participants develop a deeper understanding of mutual needs and the overall requirements of each experiment. A higher degree of coordination between analysts and FBE planners will result, and coordination would actually be eased through a deeper mutual understanding of the FBE experiment system.

The perspective taken here is analyses within the context of large systems of very complex technologies, interactions, and processes. It is possible (in fact nearly certain) that perspectives will differ depending on where an individual resides within the experiment system. What is offered here is a conceptual way in which these differing views may work together, retaining their differences, while adding to success of the whole. The structure outlined here is not new, it already exists. What is done is to identify the components of the structure in such a way that a process that leads to quality, coordinated planning and analysis can be implemented.

FBEs as a Multi-Layered System of Processes:
All aspects of this system, from concepts development through hardware systems utilization, are multi-layered. This implies that, at the highest level or layer, one is dealing with highly aggregated concepts and systems. As one goes to lower levels, detail increases, and the amount of information needed to represent or interact with a system increases greatly, often forcing one to deal with a single sub-system at a time.

For FBE experimentation and analysis we suggest that a useful delineation is four conceptual layers: missions, operations methods, systems solutions, and hardware solutions. The following section on concepts provides explanation of these four levels.

From this analysis perspective, in planning analysis of a concept, it is important to ensure that all of the process steps,

concept statement–>conceptual MOEs–>data capture–>
data processing–>analysis–>MOEs–>reporting,

are well structured, which enables both experimental objectives and analysis planning. Well structured means that one understands the levels of concepts, their sub-concepts, and ensures

appropriate corresponding levels for the other parts of the process. The importance of this cannot be overstressed since the level at which one is examining a concept strongly influences the data and analysis plans. E.g. if one is examining the Ring of Fire concept at the tactical level, it will be necessary to obtain data that identifies latencies in the C2 system, as well as many other measures that deal with the degree of success of a particular call for fires. Thus, one must insure that critical communication nodes and decision processes are sampled. Testing this same concept at another level will require a different data capture plan.

<u>Examination of Concepts</u>:
For our purposes a "concept" is defined to be:

A description of a desired result of a military operation, or an operational method, which can be tested through experimentation.

This sounds simplistic, but the inclusion of "which can be tested" is demanding. The concept statement must enable a means of testing its validity or of evolving the concept into something operationally useful. The methodology used for testing the concept will depend on its level, e.g. a high level concept may have no specific MOEs and be tested/explored through a seminar game, while a hardware concept could involve experimentation or simulation to evaluate specific MOEs associated with its effectiveness in a tactical situation.

There are several levels, or grouping of concepts. Understanding a particular concept's level assists in developing an experimentation plan. The following four levels are suggested and some concepts to illustrate each are presented (with no associated concept statements). There is nothing sacrosanct about this particular grouping. We use it only to illustrate the type of hierarchical segmentation that is needed.

Mission Area	**Operations Methods**	**System Solutions**	**Hardware Solutions**
MCM	**OMFTS & STOM**	**Network Centric**	**AAV**
Strike	Ring of Fire	COP	Hydra-7
Fleet Self Defense	Organic MCM	LAWS	THAAD
Counter Area Denial	CJTF	COMPASS	HPAC
Ground Force Insertion	Land Force Cmdr.	VIC	Mark 54
Airspace Dominance	Air Expedit. Force	Anchor Desk	ERGM
TBMD	Linebacker		Microbots

A general example of concept driven experimentation will be useful here. Consider the much agonized over concept of a Common Operating Picture (COP). A testable concept could be that a particular information set will lead to COP improvements which will improve execution of a particular mission. We now know the level of the concept through the scenario and specifics about the information being processed. We now decide whether we wish to test the information pipelines and displays or the decision processes, or both, which determines the data that needs to be captured. At the end of the experiment we can make subjective statements about the degree of success of the concept from evaluative statements from the operators. If the appropriate data

was captured, we can analytically determine MOEs, such as latencies and number of tracks processed. The final results can range from

the concept doesn't work and should be abandoned, to

the following difficulties were encountered, these specific improvements should be made, and the concept tested again, to

the concept is mature, and should be implemented or an acquisition program started.

Scenarios:
Concept testing is done within operational scenarios. Scenarios 1) set up an operational situation appropriate for the concept being tested, 2) establish a framework for expression of results (context), providing an understandable operational meaning, and 3) provide a means to differentiate the causal reasons for results variability as caused by changes in system components, OPSIT, procedures, etc. A rational cause-and-effect process would not be possible without well-defined scenarios.

Scenario evolution is a natural part of concept evolution, responding to new operations understanding, systems introduction, etc. Evolution can be incremental so that results from one scenario can be mapped into another, thereby allowing identification of improvements causes.

Using unrelated scenarios for each study makes it difficult to relate results from one event to another. Also, there are savings of effort by using the same, or closely related, scenarios for multiple studies. At the present time, we concentrate on the following scenarios:

FBEs   Culebra   DD21-DRM   Global

Efforts are ongoing within the experimentation community to identify commonality between the scenarios used for various activities, and even to modify scenarios to insure a higher degree of commonality. This will this will greatly facilitate relating results and a logical progression of concept evolution.

Robust MOEs:
The need for Measures of Effectiveness if one is to produce useful assessments is well understood. These measures have the following attributes:

a. enable conclusions about the concept being tested

b. warranted by assessment commissioners as valid measures

c. information gathered during the event allows their determination.

These simple attributes are often not met, leading to inconclusive results. Attribute (c) places demands on MOE development and experimentation design. It is possible to develop reasonable MOEs that cannot be determined from the data available during an event. MOE development should be iterative, with the data capture plan being checked to insure proper data will be available, and either the data plan or MOE being modified, if necessary.

MOEs should be robust, valid for more than a single event or concept. Robustness allows tracking of results from one event to the next, enabling concept evolution and comparison of concepts. If MOEs are changed with time, it is important to record the changes made so that the relationship between past and present MOEs with respect to concept conclusions can be made. Robustness and broad applicability also make data capture and analysis easier.

There is sometimes confusion concerning various types of measures, specifically MOPs, MOEs, and MOFEs. Measures of Performance are normally used to assess how well a system performs a design function, e.g. a processing time. They are sometimes confused with a Measure of Effectiveness, which addresses how effectively a system, or normally a system-of-systems, performs a higher level function. An example would be how effective a particular TCT process, including the hardware, humans, and TTPs are is in prosecuting targets. Measures of Force Effectiveness are very difficult to determine, or even to define. Of course, it deals with capabilities to carry out an operational mission, and requires knowledge of both own and opposition force objectives.

Note that, in an above listing of process steps, we show conceptual MOEs in the initial stages of the process and also MOEs at the end. This is in recognition that proposed measures are an important part of the design process, but that the way results are best reported may not emerge until the final stage of results production.

Successive Experimentation and Iteration:
There is a tendency to treat major experiments as independent events which produce final results for a specific set of questions. But, FBEs concepts are broad and require a succession of experiments before obtaining final answers. We expect experimentation to lead to modification of many aspects of the operations concepts being tested over time. Concepts, procedures, systems, etc., will all be in evolution. This makes it important to have an analysis system which is robust and complete in MOEs, parameters produced, and information archiving.

As has been noted above, all aspects of FBEs are multi-layered. This is also true of iteration. One must be prepared to iterate at many levels, for example:

Iteration between concepts, MOEs, and the data capture plan will be required to ensure that requirements for an experiment will be met.

At the end of an experiment, the results will be used to modify concepts, and concept iteration will affect future experiments.

The fleet involved in an experiment will have a set of exercise goals, which will be one of the drivers in defining an FBE. FBE results will modify operating procedures, which will lead to new Fleet requirements for experimentation.

Advances in understanding procedures and systems needed to be successful with a concept will lead to new operational possibilities and scenarios iteration.

One must insure that the FBE design and analysis process supports all of the needed feedback loops. Thus, results must be presented in such a way that they support multi-layered evolution, concepts, CONOPS, TTPs, doctrine, etc.

X. Knowledge Management Structure

INTRODUCTION

The term "Knowledge Management" is used here in the broadest sense. We refer to managing numerical values obtained from an automated collection system, human subjective opinions, synthesis results, results tailored to address specific long-range initiatives, etc. The challenge is to create a knowledge management, KM, system that enables archival, retrieval, and analysis. This paper describes a KM system that supports the analysis of military capabilities.

The information to be archived in this system come from differing sources: studies, wargames, and field experiments. A characteristic of these events is their variability. They have neither a common structure nor a common core of assumptions. In fact, there is an overt desire to test a range of operational structures and situations so that even a given type of event, such as Fleet Battle Experiments, will have some of its conditions change from event to event. On the other hand, there is a desire to synthesize results from many events to obtain conclusions on current and future operational capabilities. This means that the KM system must be robust to changes in the configurations under which information is obtained and developed.

The KM system must it support the strategic, operational, and tactical questions being addressed, or that may later be addressed, by the events. The system described here has been created specifically for Fleet Battle Experiments (FBEs). Fortunately, FBEs are very broad in configuration and the issues they have supported development of the required KM system.

Within DoD, there are a relatively small number of overarching concepts under consideration. For a high-level organizing structure, we use concepts such as:

        Joint Maritime Access Control (JMAC)
        Time Critical Targeting (TCT)
        Theater Ballistic Missile Defense (TBMD)
        Full Dimension Protection (FDP)
        Mine Counter Measures (MCM)
        Network Centric Warfare (NCW)
        Common Operations Picture (COP)

These are an illustration, not an all-inclusive list. They are important operational concepts that support multi-level questions, including high-level questions such as "should we operate in the littoral?", "can we support widely dispersed troops ashore?".

The above concepts are not independent nor are they of the same type. JMAC is a strategic goal and TCT, TBMD, FDP, and MCM are operations needed in support of that goal. NCW is an information superiority concept which can aid or enable operations rather than a type of operation. COP is a needed tool within NCW. Even though there are structural differences between these "concepts", they are often treated as being at the same level when planning a complex event. This is not a problem as long as one recognizes the differences and sets up a KM structure which allows the proper relationships between them when analyzing the events.

Developing a KM system requires that there be an archiving methodology which supports the "thread pulling" method used for developing and retrieving information. When archiving we place several appropriate "tags" on each piece of data. (We are using the term data very broadly here). Information is retrieved by "pulling" on a set of tags, which we refer to as thread analysis. Thread analysis starts with a specific question, from which a set of tags is defined to pull the appropriate data. The data archive must be as robust as possible with respect to thread analysis for a wide range of questions, i.e., the system must be set up so that one can access every piece of data that has applicability to a given area of inquiry. This requires an extensive set of tags and several tags on each datum. If the tagging system is not set up wisely, the number of tags needed on a particular datum can get out of hand. We have found that the answer to this dilemma of the need to balance robustness and cumbersome tagging is to take an object oriented approach, which is described later in this document.

Knowledge, information, data, regardless of the semantics used, occur in a hierarchy or at levels. There are no hard and fast rules for the number of levels and how they are defined. The number depends on the granularity desired for information. The definition depends on the specifics of the system being examined. In general, too many levels (fine grained) leads to an overly complex KM system which is arduous and time consuming to use. For our purposes, we prefer three levels. They are:

Level-1 - objective and subjective data that directly address events (event data).

Level-2 - conclusions concerning the performance of a system (system information).

Level-3 - conclusions that address capabilities at the initiative level (results).

Note that we are now discriminating between the terms data, information, and results.

Level-1 data consists of event descriptions and the time at which events occur. Data can be obtained from an automated acquisition system or from an observer recording an occurrence. Data also includes observations of the status of systems, work-arounds, configuration changes, etc. that occur at a particular time.

Level-2 information often is a subjective opinion, but it can also be a conclusion developed from Level-1 data. There is no time associated with them but they should be in the "context" of a particular operation, platform, command and control configuration, etc. These contexts can change with time. Context information is often referred to as "meta-data". As noted above, Level-2 data refers to systems. "System" is not meant to apply only to hardware. It often will refer to a C2 system, and can also refer to a process. The only requirements for something to be called a system are that it be an identifiable entity and that the interrelationships between its components can be defined. One must also be able to identify the interactions between the system and its external world.

Level-3 results will most often be pulled from Levels 1 and 2 through thread analysis. Expert opinion may also be directly inputted to the KM system database. When this is done, developing

supporting information from Levels 1 and 2 should be done or the validity of the results may be suspect.

It is important to recognize that questions have levels in order to couple questions to the proper thread-pulling analysis from the KM system. Question levels are not the same as data levels, and the answer to a given question will usually require accessing more than one KM level. Examples most easily illustrate this. Consider two questions:

1. Does a particular system (or method) shorten the TCT time line.

2. Does a particular COP configuration aid in reducing the TCT time line.

The first question is straightforward, and the answer requires pulling the appropriate Level-1 data, in particular the times required to perform the various TCT processes. One requirement is that there be a performance baseline from which the comparison can be made if sense is to be made of "shorten". The second question is more complex. Answering it can require that one pull data and information from Levels-1 and -2, then attempt to isolate the influence of the COP from other factors. The answers to these questions can provide KM system information and results at Levels-2 and-3.

QUESTION, DATA-LEVEL, AND SYSTEM RELATIONSHIPS

The purpose of the KM system is to support analysis of operational capabilities through the examination of processes and systems. Questions form the basis for analysis and are the key to reaching into the KM system. Information may be desired about a system that provides an end-to-end capability, or one of its subsystems. The question could concern the effectiveness of task performance, perhaps using an MOE such as time, or it could be the value of a particular parameter, such as reliability. One must devise the three-level KM system, and the associated data tags, so that a wide diversity of questions can be supported.

The relationships between and within KM system levels are important. A systematic methodology is needed to aggregate data in Level-1 into information for Level-2. There must be coherence between opinions inserted directly into Level-2 and the information pulled from Level-1. The same is true for the relationships between Level-2 information and Level-3 results.

Tags placed on information and data are the keys to accessing them. The analysis methodology that allows one to go from a question to the correct set of keys to obtain the desired answer must be logical, reasonably transparent, and fairly easy to use. The basic requirements for a viable data tagging scheme are that:

> there be an easy correspondence between questions
> addressed to a particular KM level and the tags, and

> there be a logical relationship between the tags for the three levels.

The relationships between the types of questions and the three KM Levels are fairly straightforward when one refers to the definition of the Levels given above. Where the various types of system information can be found and how that relates to questions is more complex. To illustrate we consider possible questions that address the two Levels. First, two Level-1 system questions:

    a. How long does it take from detection of a time critical target to the time when weapon/target pairing is completed?

    b. How long does it take to perform target mensuration?

Both are Level-1 questions because they refer to the value of a particular system parameter, time for both questions. The first time can be obtained by summing the processing times for the appropriate parts of the total system. It may be necessary to follow sets of events through the system to obtain the times, or the processing times may be archived. The second time is obtained by pulling data for the subsystem that performs mensuration.

Second, consider two Level-2 questions:

    a. Does incorporation of an ISR desk to manage sensors improve the TCT process?

    b. Does an ISR desk improve the quality of forwarded target folders?

Both are Level-2 questions because they do not ask for a parameter or MOE, but for information about system performance. The first question concerns a macro-process, TCT, and the second for a sub-process, target folders generation.

It may be that the answers to the Level-2 questions can be obtained by reaching into only information in Level-2 or it may be necessary to also pull out some Level-1 data. Note that both questions contain the word "improve". This implies that a comparison is needed between performing a process with and without and ISR desk, which means that somewhere there must be baseline, or without the desk, information. This means that information, or tags, must be present in the KM system that identifies the configurations that were in use when data/information were collected. The above examples illustrate that a fair amount of care is needed to relate questions and the tags.

The fact that Level-2 data will contain subjective system information implies that the information will not make sense unless one has a good definition of what the system is. This is true, and especially important for C2I systems because they are in a near-continuous state of evolution. Thus, maps of the various C2I systems are required as supporting information for the data archive (accessed through tags on the data and information). In addition, there are many hardware and software differences between experiments, so that supporting configuration information and tags must also be included.

DATA TAGGING STRUCTURE

The tagging structure must map in a fairly transparent way on the objects and events that make up military operations. The number of categories should be small to reduce complexity, and there should be no overlapping categories, which would create uncertainty in how to tag and make information retrieval difficult. These criteria can be accomplished by defining three categories:

> Things – objects, systems, or people that perform actions.

> Attributes – descriptions of the state or characteristics of things.

> Actions – activities that occurs at a particular time that change
> the state of a thing or are interactions between things.

There are subtleties involved in using a simplified tagging structure of this type. For example, suppose the piece of information to be archived is an action taken by an object. The obvious tags are those that identify the object and the action taken. In addition, it is probably important to know the attributes of the object to place the information in the proper context. Thus tags from all three categories can be needed for the datum. Almost never will data be tagged with only one of the above categories.

Within these three categories it is possible to identify the sets of Things, Attributes, and Actions that will adequately describe military operations. They are:

Things	Attributes	Actions
Platform	Status	Transmit
Sensor	Mission	Receive
Weapon	Location	Detect
Information	Command Relation	Decide
C2 System	Performance	Command
Assessment Node	Workload	Physical
Planning Node	Capacity	
Command Node		
Force		
Person		

Some of these are obvious, some not, and examination of the lists could lead one to conclude some are downright strange. A full description of the underlying rationale is beyond the scope of this paper, but a few examples to illustrate the basic ideas are warranted.

> C2 system refers to the full system, whereas Command Node is
> an activity that issues commands, be it a single person or CIC.

> Physical refers to any physical action, such as fire or reposition.

> Command is the issuance of a command at its source. Transmit

refers to sending any information, including commands.

A force is any size grouping of platforms.

Workload refers to how much activity a thing has to perform and capacity is the current ability to carry out its activity. Capacity can also be physical, such as how many rounds in a magazine.

Obviously, there are many sub-tags within each of these descriptors. There are types and identifiers, e.g. Platform includes a tag for the type, such as ship or airplane and the identification of the specific platform. Decide could be a decision to engage a target or it could be a decision made for BDA assessment. Data and information will have tags that identify where they fit within these categories. A given piece of data will have more than one tag, e.g. to tag sensor information being sent: sensor, platform, sensor type, location, sensor information, transmission.

SYSTEM DEFINITION

The purpose of this section is not to define the word system, but to indicate how one links a question or thread analysis to what one considers to be the system for the particular case. In much of what follows in this paper we use "sensor system" as the example, broadly defined to be all components and actions from the point at which a target is detected to the point at which weapon/target pairing is accomplished. With this definition, one can list those functions performed by this system:

Sensor System Functions (Actions)

At Sensor Platform	At ISR Center
Receive Command	Receive Data
Move Platform	Fuse Data
Command Sensor	Interpret Data / Decide
Search	Assign Sensor
Detect	Create Target Folder
Transmit Data	Send Folder
	Mensurate
	Nominate Target and Transmit

The above assumes that there is some type of central function, ISR Center, or more than one, that receives sensor information and acts on it. However, the listed functions are independent of the exact structure of the sensor system. The functions are shown in the order that data is developed by the system, starting with the sensor on a platform being moved into position, through detection and transmitting information, processing the information at an ISR site, and sending the final target nomination out of the system.

Having defined the system, it is fairly easy to see the tags that would be attached to data for one of these functions. There would be significant number of meta-tags that identify the experiment, platform, and other context. Then there is the Sensor tag to identify the data as belonging to the

sensor class, followed by other tags to designate specifics, such as it refers to transmission of a command to the sensor platform from the ISR desk.

A datum will have only one set of meta-tags, but it can have more than one set of system tags. For example, information probably will be associated with more than one system, such as the sensor system and the COP. Thus, tags appropriate to both will be present. This allows inter-system relationships to be examined. An example could be how a specific sensor control configuration contributes to an improved COP.

LEVEL-3, RESULTS AND ANALYSIS

Level-3 results address capabilities at the operational level. Thus the tags will be derived from major operational concepts, such as:

TCT	Time Critical Targets
STOM	Ship to Objective Maneuver
NCW	Network Centric Warfare
COP	Common Operating Picture
CAS	Close Air Support
Etc.	

These results will have been obtained from a single experimentation event or by synthesis of results from several events. In either case, the result in the database needs to have a tag that identifies the event(s). It is not only possible, but probable that information that applies to one concept can apply to one or more others. Thus the result will have tags for each of the concepts to which it applies.

Many of the results will have been developed from information in Level-2 or even from data in Level-1. It is important to identify the trail(s) through the data from which the result was developed. Tags are also used for this purpose. This allows one to access the supporting evidence for a result.

The best way to understand the relationships between analysis, the data system structure, and the use of tags is to consider an example analysis question. The following is a constructed, rudimentary example of the process, presented as a set of logical steps. It illustrates a thread.

Analysis Question: How well can we do TCT? (Note poorly constructed, broad question)

Results Pull: Pull TCT tagged data from Level-3.
      Results: A is good.
                B is not so good.
                Concurrent FDP and TCT with the same platform is difficult .
                Etc.

These pulled results may be sufficient as-is or one may wish to use them as a starting point to explore more deeply. Then, one needs to ask a more in-depth, specific question and do another

information pull, probably from Level-2 and even Level-1. Continuing with this example, assume the interest is in the concurrent FDP/TCT result.

Analysis Question: What interaction between TCT and FDP reduces the ability to do TCT?

Information Pull: Pull TCT and FDP tagged information from Level-2.
                        Only pull information that has the same platform tag.
    Information: Difficulties with tube loading
                        Insufficient SAM rounds
                        TCT C2 configuration too slow for FDP

This information focuses one on one or several aspects of the problem. At this point a third (or more) analysis questions can be posed. In this way the thread of information is built up.

3rd Pulls: Pull the connected C2, weapon system, sensor system:
                        information from Level-2, and
                        data from Level-1.

Other iterations in the process will occur until the analyst is satisfied with the information or there is nothing new to be found. A result of this analysis may be to create new results and information, and archive them with the appropriate tags.

The above example focuses on using the database for analysis, starting with results that are already in Level-3. In order to have results in Level-3, analyses may have already been done. It is also possible that the results were inserted directly from expert observations made during an experiment. This introduces the need for two types of tags for Level-3 results. If the result has been inserted, the tags will identify the concept and whatever context is needed. If the result comes from analysis, it is necessary to identify the thread, for which a tag is needed.

The above analysis example started with an analysis question, accessing a result that already existed, then drilling down into Levels 2 and 1. Because of the result accessed, the drilling down began with looking for instances of TCT and FTP on the same platform. The example ignored the fact that the result was already present, and that thread paths already existed, in order to illustrate the analysis process.

LEVEL-2, SYSTEM INFORMATION TAGGING

Analysis begins with a question, then assembling the appropriate tags to pull the thread. Assuming that one wishes to generate new results, the pull will be from Levels-2 and 1. The following two sections describe the tagging schemes for these two levels.

Level-2 will contain much context meta-data. Examples are:
        Event identifier
        Operation or MESL within the event
        Type of operation being examined

Description of the specific C3I structure
Descriptions of the specific hardware and software systems
With each datum in Level-2 there will be tags identifying the associated meta-data.

Level-2 information deals with system capabilities. A example of defining a system was presented above using generic sensor as the example. It illustrates the many subsystems that make up the total system. Level-2 information can be for a system, subsystem, or combination of subsystems, as illustrated in the following diagram. Also illustrated is that archived information can be a subjective "opinion" or can be a "result" pulled from Level-1 data.

```
|----------------Level- 2 Opinion------------------------|
|                                                        |
|--L-2 Opinion-|                 |------L-2 Result---------|
|              |                 |                        |
| Sub-Syst A   | Sub-Syst B  |-----| Sub-Syst E  |
          |      |
        Sub-System
        Interaction
        Result
```
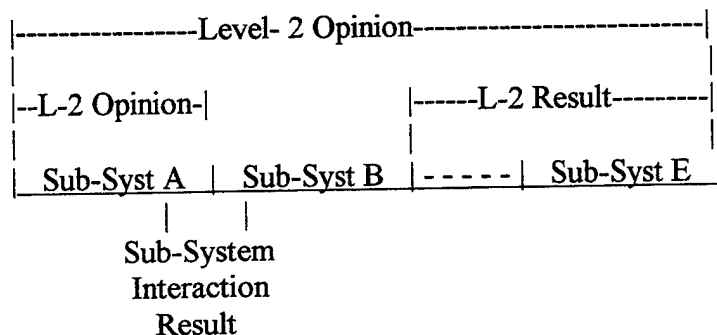
Recall that the data tagging structure has three categories: Things, Attributes, and Actions. For Level-2 information, actually for all data, the Things category tagging is natural. It consists of identifying the system itself and those things on which it sits. Attributes is also natural. The information can be a status report made at a particular time, what the mission is, the workflow, etc., and it can include more than one attribute in a single data entry. Actions at this level are more subtle as a category. At Level-2, Actions refers to information about system performance when an Action is being performed. Reporting on the status of a communication system might be that it is down. Reporting on its Action might be that the data rate is too slow for a particular peak load. Such information can be time marked, can refer to a time period, or may have no time associated with it being a general capability comment.

LEVEL-1, DATA TAGGING

The distinctive characteristic of Level-1 data is that it contains events that occur at a specific time. Event data can be subjective or objective. Examples are:

Objective: a target folder being sent to the fires cell, or
a STOW simulation target inject,

Subjective: an observation or an opinion, such as
an assessment node becoming overloaded.

Subjective opinions are needed in Level-1. An example shows the importance of doing so. Take the case of an observation that an assessment node is overloaded. There may be available for that time period objective data that three sensor hits arrived at the node within a five minute

period. Combining the subjective and objective data allows one to draw the conclusion that this node becomes overloaded if more than two targets are to be processed within a 10 min time period. This conclusion could be a Level-2 datum entry for the system.

There is little difference in the tagging for Levels-1 and –2. Event time is unique to Level-1, and Level-1 will always deal with an action, and be so tagged, while most information in Level-2 does not.

A better understanding of tagging at the two lower Levels of data/information can be obtained through examples. The sensor system is again the example. Thus, the "sensor system" tag is understood to be attached. We only list the specific tags for that data, not all the context tags, such as platform and sensor type.

Data or Information	Partial Tags
Level-1 data: Time to create folder	Decide, GISRS-M Terminal
Time to mensurate	Decide, PTW+ Terminal, Target Type, Physical Environment
Target info transmission	Information, Target-Information, Transmit, E-mail
Time for weapon/target pairing	Decide, LAWS Terminal, Fires Cell configuration reference
Level-2 info: "The fires cell configuration significantly reduces the TCT timeline when compared to a baseline configuration."	TCT, Latency, COP(?), Fires Cell configuration reference, person entering opinion or analysis thread reference

This conclusion could have been produced directly by an observer or by accessing the noted Level-1 data. If it came from the Level-1 data, perhaps from that data referred to in this example, then a better Level-2 statement and tagging would be:

Level-2 info: "The use of GISRS-M, PTW+, and LAWS in a JFE Cell configuration improved the TCT timeline."	TCT, Latency, JFE Cell, LAWS, GISRS-M, PTW+ analysis thread reference

The following is a constructed example of Level-2 information at the subsystem level.

Level-2 info: "TARPS-CD imagery did not have sufficient resolution"	TARPS-CD, Detect, Target Type, TCT, Environment, Location

If the observer logged a time at which this observation was recorded, it could be possible to correlate it with Level-1 data concerning an actual target, sensor status, etc.

XI. Data/Information Requirements

Many types of data and information are required in order to analyze successfully FBEs. The types of data needed and where they are obtained are listed here. Note that some types of data are obtained from multiple sources. Also presented is some perspective on the Navy Lessons Learned system.

NEEDED DATA/INFORMATION

Information throughput
Electronic data
Communication logs

Experiment Events
Observer data logs
Operator logs
Simulation injects
Electronic data

System and sub-system performance
Expert opinions on web data collection
Post experiment interviews
Analysis of electronic data

Process capabilities
Expert opinions on web data collection
Post experiment interviews
Analysis of electronic data
Observer evaluations

Evaluation of experiment results
Post experiment interviews
Analysis

Experiment results implementations
Interviews with the Fleet

The first two types are basic data. They are records of specific events that occur at specific times. Communication logs also provide context information. The next two deal with the performance of systems and processes. This information is both collected and produced by analysis. Evaluation of experiment results is obtained from expert opinions and from in-depth analysis of all the data and information gathered during the experiment. Implementation refers to what operational forces do with the results of an experiment.

47

DATA/INFORMATION CAPTURE MECHANISMS

There are five mechanisms for obtaining data and information.
　　electronically capture the bits and bytes within the system
　　an expert observer team
　　observation inputs to the NWDC website
　　post-experiment interviews with operators
　　post-experiment workshops
　　analysis
Analysis will not be discussed here. It is included only to indicate that some levels of information are developed through analysis of lower-level data and information. This is covered some in the next section on reporting.

Electronic Data
Electronic data must be available if one is to quantify, and even unravel, C3I timelines. One cannot develop numerical MOEs without them. For those who believe this can be done, the awful truth is revealed when opinion is compared to LAWS data. By electronic data is meant the information flow through the various electronic hardware systems, such as:

LAWS	PTW	COP
GISRS	GCCS	etc.
STOW	JCSE	

Within each of these systems there are processes. We need to capture ALL of the processes that are part of creating decisions. For example, mensuration often requires use of reference imagery. When the operator accesses the reference image library, that is an event that needs to be recorded. We need not only the time that particular type of event occurred, but also what type of data was pulled. Of course, we need all information injects from STOW.

Website
NWDC has developed a website to collect in close to real-time observations from the operators in the experiment. The information is over a wide range of topics, such as:
　　specific system performance
　　need for enhanced communication means
　　process performance and improvements suggestions
　　work-arounds to achieve success
　　usefulness of situational awareness displays
　　etc.

All of the information provided is colored by human opinion. This does not mean that it is invalid or of low value. It is of high value because an FBE is a human-in-the-loop experiment and human opinions as they carry out the operation are an important component of the overall system. Even so, one must synthesize these opinions with each other and with data to check their validity and to provide context.

Post-Experiment Interviews

Interviews with operators are the same type of information as obtained on the website. However, the provide an integration across the whole of the experiment. The perspective is different than obtained from the on-the-spot opinions. Also, the interviews are carefully structured to obtain information that applies to the learning objectives of the experiment.

Post Experiment Workshop

After the experiment, workshops are held with Experiment Leads and system operators to obtain their perspective on the experiment. This generates a great deal of valuable context information. It is also the first synthesis step in the analysis process. Special circumstances are revealed, such as a system that was performing abnormally during a particular time, or that a particular process used had a particular effect. Those who operated systems such as LAWS and GISRS have a unique perspective.

Navy Lessons Learned

A mandated part of Fleet exercises is to report into the NLL system. The data needed for FBE analysis cannot be obtained from NLL as it is currently configured. There is no chance to modify NLL to provide what is needed, and that it is not even desirable to do so. Thus, soas to not cause extra effort for the operators, NLL input should be provided by extraction from the experiment data collection. The best chance to do this is with the website collection.

Expert Observer Team

A 30-40 person expert observer team is fielded for each FBE. Their purpose is to gather event data at pre-specified sites that are important nodes in the C3I system. Operator logs and chat records provide additional information of the same type.

In addition to the simple occurrence of an event, the observer also provide context. For example, the observation that an operator was overloaded at the time of the event. Such context is valuable to unraveling cause-and-effect relationships.

The following are the directions and an example of the data logging sheets that are used by the observers. At the end of each day, the data and information on these sheets are inputted to Excell spreadsheets and transmitted to a central data archive.

EXPERT OBSERVER FORMS AND DIRECTIONS

Event Data Logging Sheet Directions

Purpose: The purpose of these sheets is to make it easier to record your observations. All you need to do is record your observation or opinion, the time, and perhaps check a couple of boxes that identify the type of observation you are making.

Data: The data is your recorded observation and the time. All other things that are checked on the sheet are context for your data. Of course, this context is quite important so the more complete you are the more useful your data will be.

Context: The following diagram explains the things we wish to have recorded.
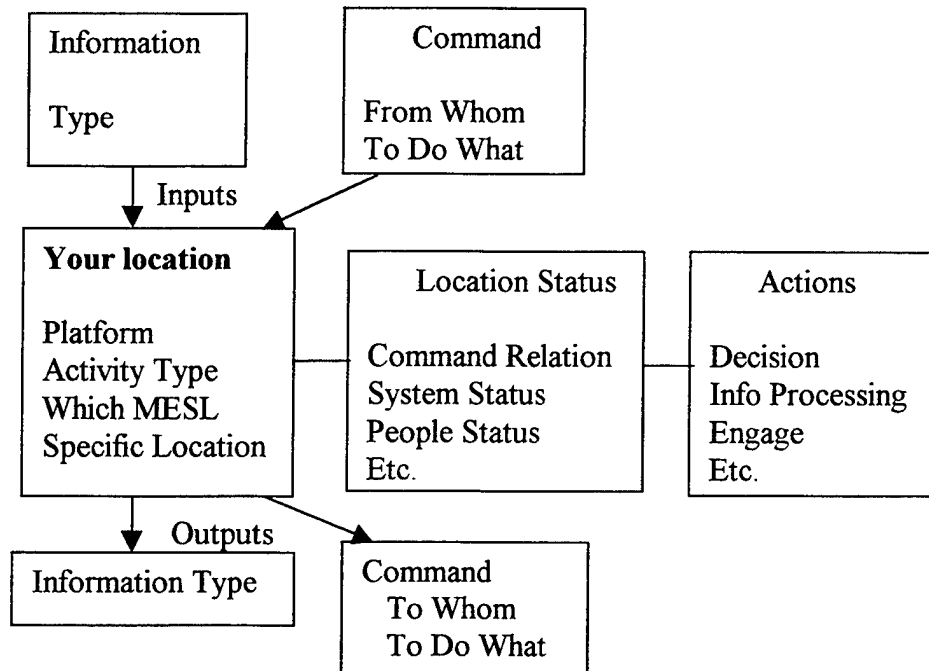
**This is not meant to imply that only those things listed
are to be recorded. They are examples. You record
those things your experience tells you are important.**

The above is quite simplistic, but with such information from many locations cause and effect can be traced through the system

Data Sheets
The following data sheet is to be used to record you observations. The basic data are your observations. BE COMPLETE. RECORD ANYTHING YOU THINK MIGHT BE PERTINENT TO UNRAVELING WHAT IS HAPPENING. The meaning of the various entries follows:

Data Logger: Your name

Specific Location: The platform you are on and the observation location on that platform.

MSEL: Identify the operation that is occurring.

Time: This is the time of the observation/event, not the time the sheet is filled out.

Your Data/Observation: This is the **data** you wish to enter
Track ID: In many cases you will be recording events that are associated with targets. Be sure to capture and record the target ID.

The remaining columns on the sheet are simple check blocks. You check a block to indicate the type of data you are entering. This is an aid to archiving data in the knowledge management system. The checks will ultimately become tags for the data in that system. If you don't know which block is appropriate, don't check any. Note that there is an other column. Feel free to invent any tag you feel is appropriate.

Information: Many events deal with information and we need to track what is happening with information processing and dissemination. In your observation column you will identify the type of information.
    Dev: The person at the observation location is developing or interpreting information.
    Rec: Information is being received (in your observation also record from where, if possible)
    Sent: Information is being sent to another location (record to where, if possible)

Command: Commands are a specific type of information and we need to identify them. You will record in the observation column what the command is.
    Sent: Command is sent (record to where, if possible)
    Received: Command received (record from where, if possible)

Status: Often systems, people, or processes are not working or have some special status. The data here is the status and which system.
    Phys: A physical system such as a terminal or communications.
    Peop: The person(s) at your observation point, such as overloaded, tired, comprehending, etc.
    Cmnd: Command and control process functions.

Other: This is yours to invent. You may observe a pattern that you feel should be a category.

FBE-F Data Logging Sheet   (FBE-data-JFC)

JFC INFORMATION DATA   Data Logger_____
    Specific Location _____   MESL _____   Date _____
Comments:

TIME	YOUR DATA / OBSERVATION	Track ID	INFORMATION Dev	Rec	Sent	COMMAND Sent	Rcvd	Phys	STATUS Peop	Cmnd	Other

Please note: Your recording events is very important. Also, you are the only source of **human condition context information.** This is an important part of understanding the C3I system.

51

The following figure shows the structure for archiving the various types of data and information that are acquired in FBEs. It also shows some of the reporting that is done, which is covered more completely in the next section.
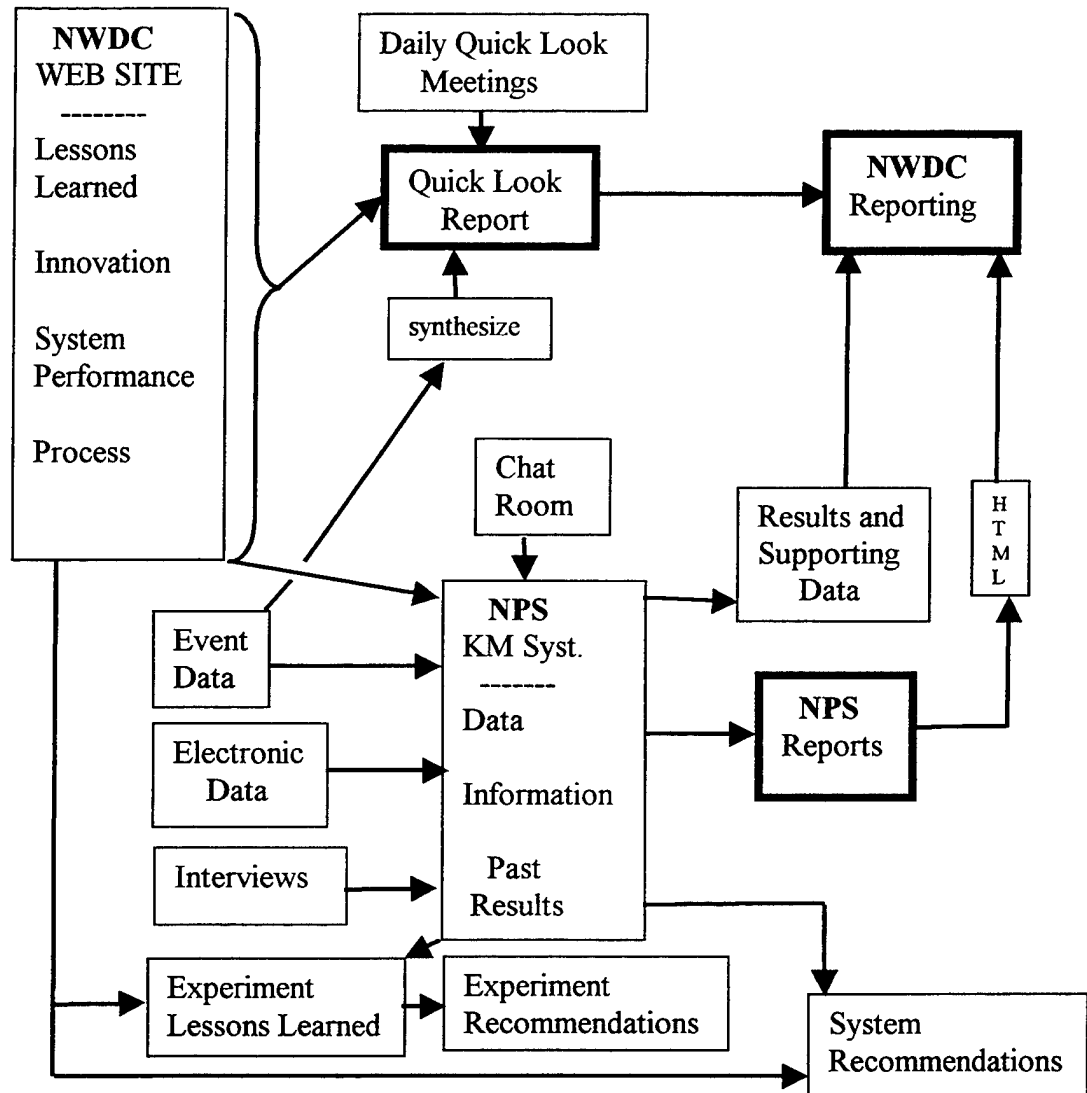
XII. Reporting Structure

There are many customers for FBE results, some internal to the process, many external. The following notes several of the customers, the types of information they need, and the reporting structure that meets those needs. The reporting structure is supported by a synergistic information structure, which is also briefly described.

The customers and the reporting venues that meet their needs follows. This list is rough and not complete; it is only meant to indicate some reporting structure.

Numbered Fleets	NWDC FBE Report
Plans and Programs Offices	Cross FBE Long-Range Analysis Reports
NWDC Direct Customers	Presentations and Briefs
MBC	IJWA FBE Report
NWDC Concepts, Doctrine	Cross FBE Long-Range Analysis Reports
J9, CHENG, BMDO	Tailored Reports
DoD world	NWDC website
	IJWA website

Underlying this reporting is the knowledge system, which has a structure that supports the above. Also, note that the two websites have corresponding structures. This is indicated in what follows. It is important to realize that putting these structures in place is an evolving process. Also, the following structures are general and do not show the complexity that has to exist in the final products.

NWDC Website Reporting System
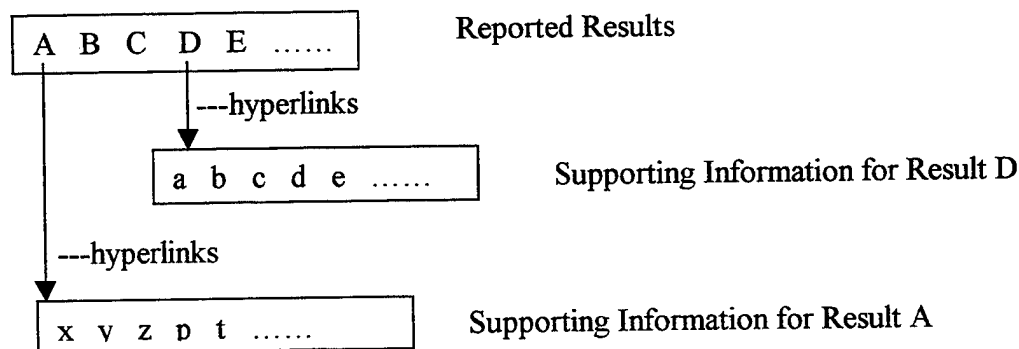
This Website provides access to a range of results through HTML links. At the top level one has synthesized results which address major concepts and initiatives. Below that is a level of supporting information so that a person who has interest can "drill down" and see the information on which a result was based. The system is deterministic, meaning that the links and

supporting information are fixed. The person browsing is not able to query a body of information to conduct his own studies.

The NPS/IJWA analysis information/data base is constructed differently. There are no predetermined links joining information and data at the various levels. Rather there are predetermined data tags so that data and information which are appropriate to a particular question or analysis can be accessed and assembled. A robust tagging scheme, not one that is focused on a particular experiment or analysis is necessary because questions will change with time as knowledge is accumulated and as needs change. Segmentation of data and information into logical types results in a three level structure, diagramed below.

IJWA Three-Level Analysis Database and System
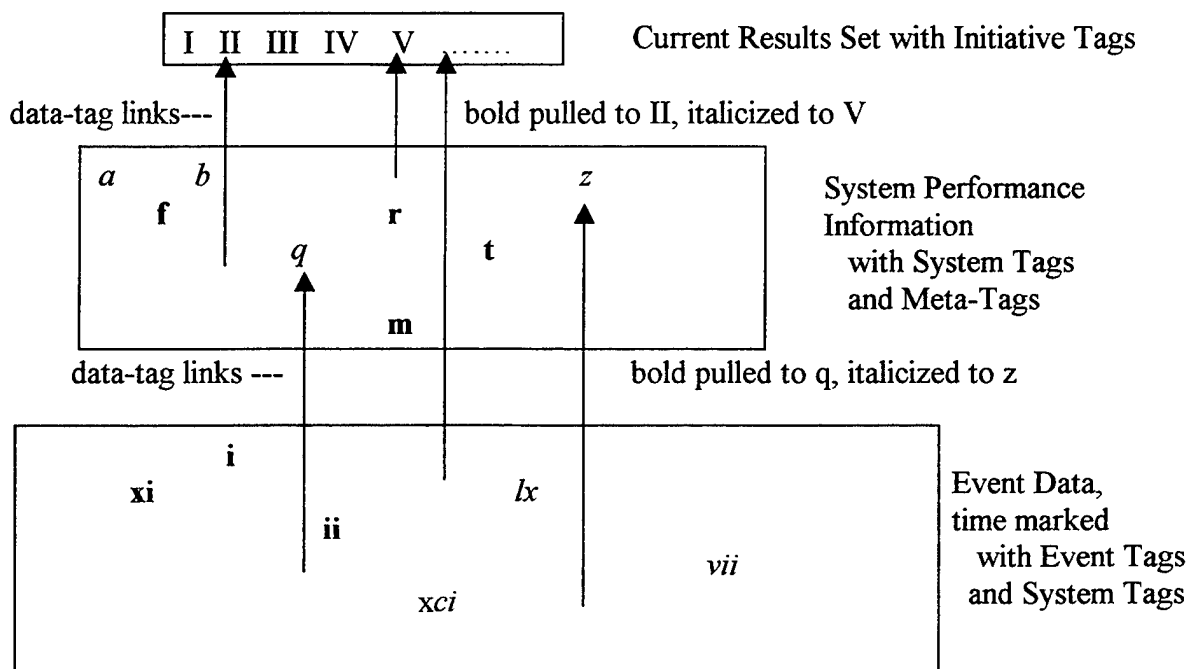
Links between results at the top level and information and data at the lower levels are not fixed. A change in a top-level question can modify the tags used, which modifies the data accessed. New data will be added to the system as new experiments and events are conducted. When this occurs it may be necessary to do a new information pull for a given question to update results.

A requirement for the analysis database is that it support the NWDC Website reporting system. There will be a direct correspondence between Website reporting and some of the KM system's top-level results. In addition, some of the information that is in the second level of the Website system will be provided by the KM system, along with the hyperlinks to access the information.

In addition to the results forwarded to the website, the KM system will contain results for specific concepts, doctrine issues, and possibly TTPs. This will necessarily be true to support all of the customers noted above. A requirement for all reporting is to insure consistency between the results reported out to the various customers and by various products.

An important question is whether there should be a direct link between the Website reporting and the KM systems. Doing so could enable a person probing the NWDC results page to reach directly into the KM database. We believe this should not be done. Reaching into the KM system properly requires extensive knowledge about the data tagging methodology, how to assemble a collection of tags to extract information concerning a specific question. It is probably beyond the scope of the results page to provide such knowledge. There is the danger that a person can use tags incorrectly and pull inappropriate or incomplete information, compromising their confidence in the displayed results.

The combination of the Website and the KM system does for us provides the following:

> The Website provides a wide audience easy access to principal results.

> The Website structure also provides easy access to information
> that supports the results.

> The KM system provides a large body of information to warrant the
> validity of reported results.

> The KM system provides a means for archiving and correlating information
> for multiple events, synthesizing results for initiative progress.

> The KM system provides allows one to correlate information whose relationships
> are not easily accessible, revealing new and perhaps unexpected factors.

> The KM system allows one to link a wide variety of inquiries to all related factors
> which must be considered in reaching a conclusion.

IJWA REPORT TO NWDC AND NWDC FINAL REPORT

Two principal reports are produced for an FBE. IJWA reports analysis results to NWDC, which then reports the official results for the experiment. The basic content and structure of these reports are:

FBE Analysis Report
        contains results from analysis of basic data and information
        is organized along systems and process lines
        structure enables extraction of results for future studies

FBE Final Report
        contains an Executive Summary of principal results
        is organized along initiative lines

The following diagram illustrates the relationships between these reports and the associated data and information. Note the correspondence between this and the levels described for the knowledge management system.
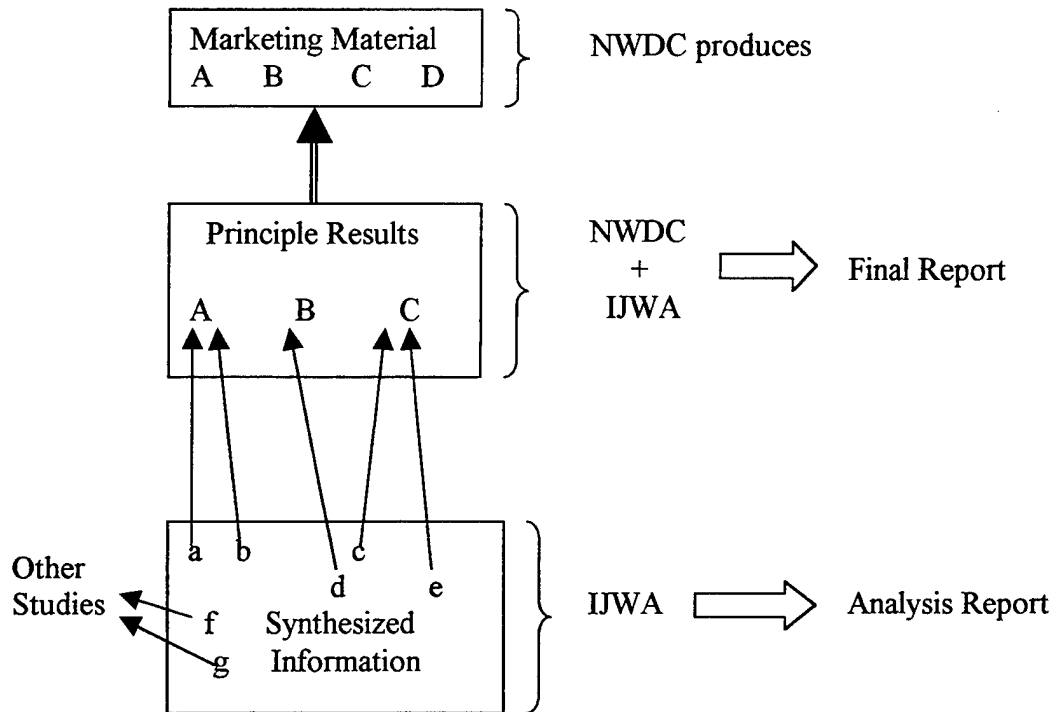
The arrows are included to indicate that information developed from the analysis process (synthesized information) is brought together to produce the final principal results. This information is also used to develop results for other studies. It is also apparent from the figure that the structures of the principal results and the synthesized information are not the same.

XIII. Case Study Analysis
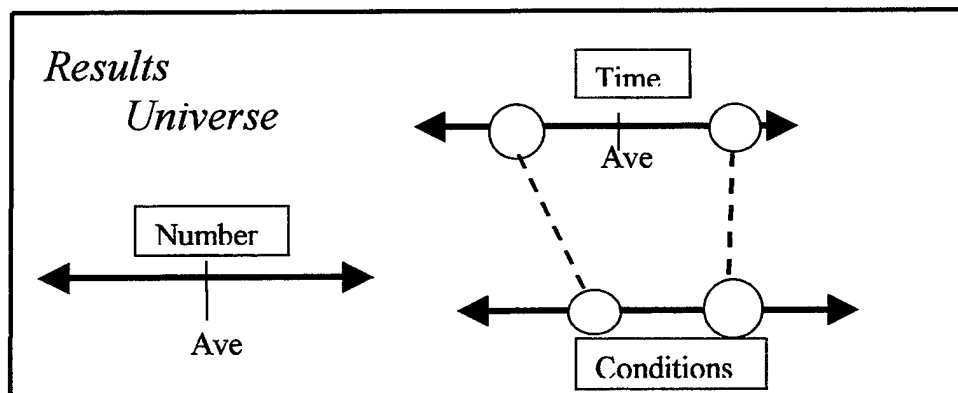
Measured data are of many types: e.g. pure numbers (or fractions), time is often an important parameter, etc. These measured values are data elements from which information is developed, including MOEs. Here, for simplicity, we consider a particular Time of interest. Related data are needed for analysis, and we refer to this simply as Conditions.

The figure shows that parameters have ranges of values, which were measured for Number and Time, and which were recorded by some means for Conditions. Number and Time data can be manipulated mathematically, such as calculating an average. Conditions are manipulated in various ways depending on what they represent, in this case grouped into categories.

Number and Time can be processed into averages, and the extremes are known. What should be reported to convey capabilities? Time extremes might be very important and worth reporting, average time probably doesn't convey much information. One could use any of these numbers as an MOE, but it is questionable whether they are measures of operational "effectiveness". If they are used, they should be related to a standard or there is little effectiveness information.

CASE STUDY: It is useful to determine and convey those Conditions for which the Time is long or short. This is the relation between Time and Conditions shown in the figure by dotted lines. A way of presenting the results is by a table. In the following table S and L refer to short and long Times and A, B, and D refer to conditions.

	T	C		Operational requirements can
	S	A		be met for conditions A.
acceptable	S	A		
times	S	A		
	L	B	process	Requirements cannot be met for
	L	B	improvement	conditions B and D and process
	L	D	needed	improvement is needed.

It may be possible from this accumulation of information to unearth cause and effect. The results may point toward changes needed to meet requirements for Conditions B and D. If changes are made, further experimentation would be needed to determine if it works.

RESULTS PRESENTATION: How should such results be presented? Consider the following parameters for reporting (not an exhaustive set):

1. Average Time	most compact	least information
2. Time Extremes	two number needed	more information
3. Fraction Meeting Requirement	need fraction and requirement	better information

Each of these is a numerical value that could be called an MOE. Calling the first two MOEs is questionable, but a case could be made for the third. In any event, it is obvious that reporting any of these numbers by itself leaves out much important information. For this example, it is clear that the most important information is Conditions.

This example leads to several important points:

A. Data elements can be (should be) specified before an experiment, the same for learning objectives, but you may not be able to define MOEs a-priori. One can imagine for this simple example that the MOE emerged from the results.

B. As information is rolled into a smaller number of result parameters information is lost, and it may be crucial information.

C. Most often a result will be a set of information. If an MOE is desired rather than the set, it should be developed, and presented, in such a way that it points to the full set of information/results that is to be conveyed.

D. Include with an MOE should be the more detailed information that tells the whole story. In the example used here, the information to be conveyed is the Conditions for which requirements can/cannot be met. Fraction Meeting Requirement is the logical choice to alert the recipient of the results to this information.

Of course, the above is referred to as a Case Study because the information conveyed is "those cases for which..."

Appendix A.

THE MODULAR COMMAND AND CONTROL EVALUATION STRUCTURE (MCES)

A. Introduction

The MCES is a general approach to evaluating C3 systems which has been successfully applied to a number of issues concerning C3 system planning, acquisition, testing and operation. It augments traditional analysis by providing a series of seven steps or modules to evaluate alternative C3 systems and architectures. These modules guide analysts who might otherwise focus prematurely on the quantitative model rather than the problem definition and the specific measures needed to discriminate between alternatives. The seven steps of the MCES are briefly described below including the product of each module.

The MCES begins by identifying the objective of a particular application. This leads to a formal problem statement. The second step is to bound the C3 system involved, by producing a complete list of system elements at several levels. The third step is building a dynamic framework that identifies the relevant C3 process—a set of functions. These are derived from the generic control loop (cybernetic) model of C3. The fourth step combines the results of steps two and three by integrating the system elements and the process functions into a model or representation of the C3 system. The product of this module is at least a complete descriptive conceptual model and sometimes a complete mathematical model. The next (fifth) step is to specifically identify measures of performance, effectiveness and force effectiveness at the corresponding levels of the C3 system and function. The sixth step is to generate results or values for these measures by testing, simulation, computational modeling or subjective evaluation. Finally, the various measures are aggregated and interpreted in the last step. Each of those steps is described as a module below.

In a new area such as C3, standard language and paradigms are difficult but necessary. The MCES was developed by a team of experts from industry, government and academia and was endorsed by the Military Operations Research Society. It presents difficult concepts in a standardized way that is easily absorbed by both new practitioners and managers. MCES has potential for reducing mis-understandings of the purpose and mis-applicability of analytical results. This is important when issues of great diversity of nature, size and level of detail are being considered, such as in preparation of the Program Objective memoranda (POM). Standardization of analytical procedure can be advantageous if based on a comprehensive and rigorous methodology such as MCES. MCES can be used for studies ranging from the quick conceptual level to the complete quantitative study. It is difficult if not impossible to require a complete quantitative study for each issue during a POM cycle, as is required for acquisition cycle issues with the Cost and Operational Effectiveness Analysis (COEA). But application of the MCES at even the conceptual level of analysis may allow better articulation of POM tradeoffs. The next section is an exposition of the substance of the MCES. This serves as preparation for the required interpretation of the MCES in terms of the MEB C3 problem as specified in Task 1. It will then be followed by application of the MCES to the allocation of SINCGARS as also required in Task 1.

The seven steps of the MCES are performed iteratively with the decision maker as shown in Figure 1. Iteration is an important concept which
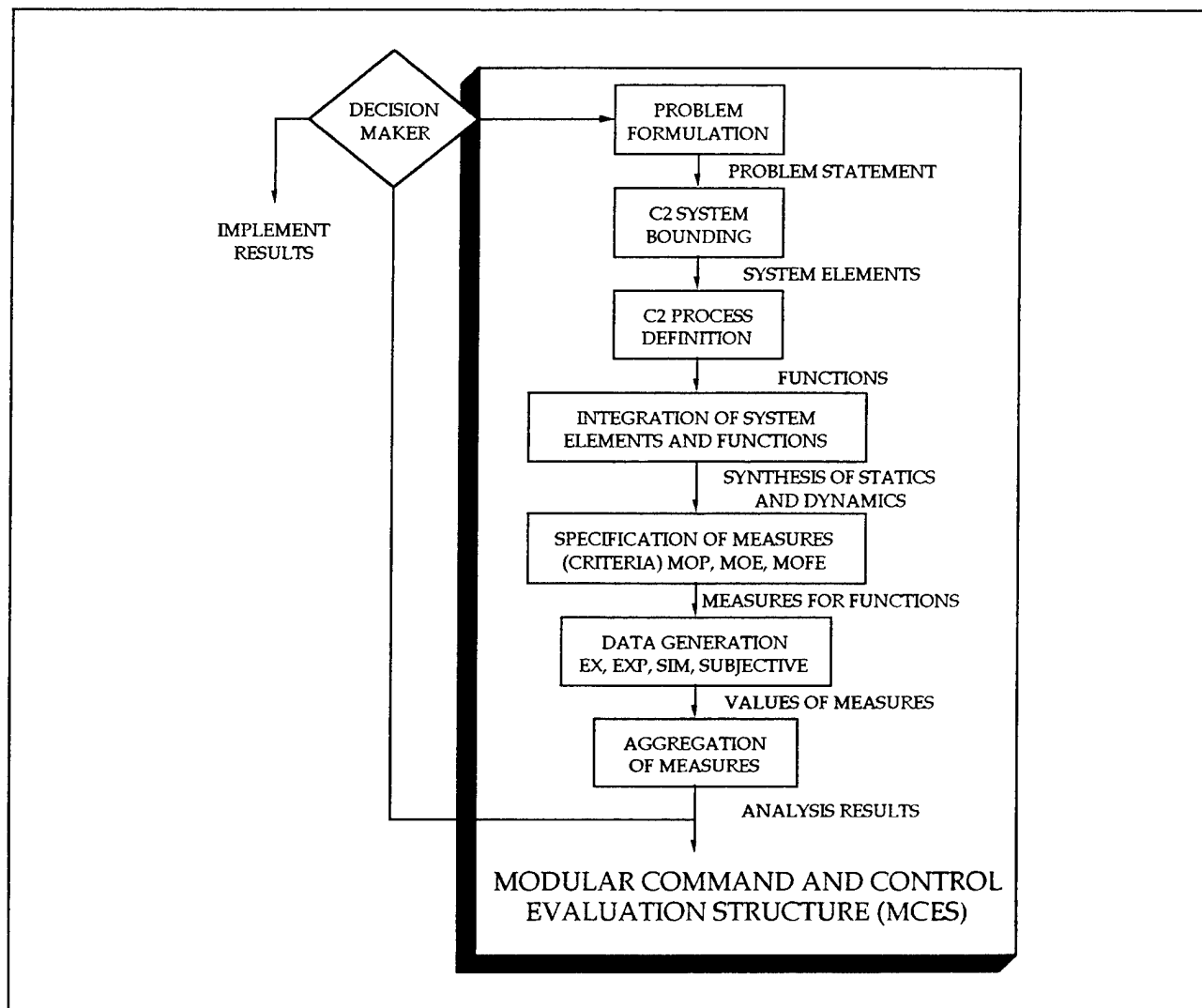
Figure 1. Modular Command and Control Evaluation Structure

prevents "paralysis by analysis." Iterative refinement of the problem and analysis helps both the decision maker and the analyst to prevent studies from being overtaken by events The outputs of each step are also shown in Figure 1. Each of the steps or modules is explained below.

B. MCES Modules

1.   Module 1: Problem Formulation

Module 1 describes the decision maker's objective and the context for a specific C3 problem as shown in Figure 2. In it the formal decision process (if any), the policy assumptions and the scope and depth of analysis are defined. The identification of the full set of decision makers being addressed may be necessary. In this module both the appropriate mission and scenario(s) are made explicit. The output, a precise statement of the problem, is used in the second module to bound the C3 system of interest.

The objectives of the decision maker(s) posing the problem are identified in terms of the life cycle of the C3 system and the level of analysis prescribed. The decision maker's objectives generally reflect the various phases of the life cycle of the C3 system, namely: (1) concept definition and/or development; (2) design; (3) acquisition; or (4) operations. The appropriate

level of analysis is derived from: (1) the mission the system is addressing; (2) the type of system itself; (3) the timing, scope and criticality of decision; and (4) the background and commitment of the decision maker(s). In this problem formulation step, it is wise to make an initial pass at all the MCES steps with the objective of identifying the range of likely answers for each module. This helps scope the analytical effort as early as possible.
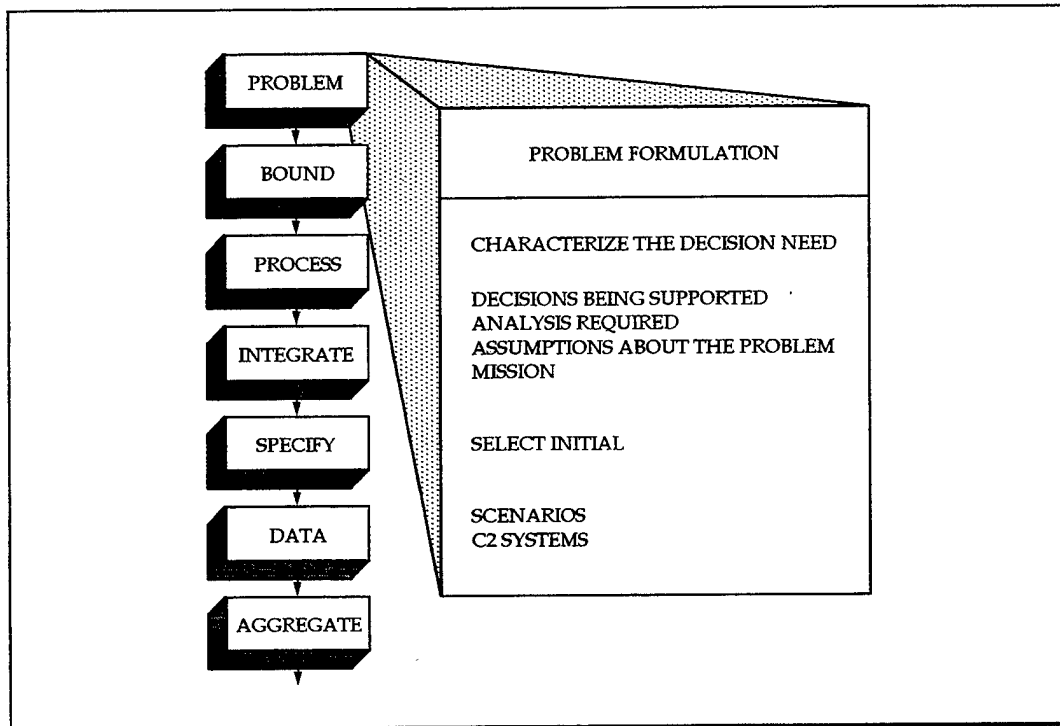
Figure 2. MCES Problem Formulation

In the implementation of this step, the answers to several questions may provide guidance, namely:

1. Who is/are the decision maker(s), and how and when will the decisions be made?

2. What mission area is involved? Must joint or combined forces be addressed?

3. What communities/viewpoints must be addressed for acceptance?

4. What are the basic assumptions of the problem? Classification level? Historically how has the problem been solved?

5. Does the evaluation apply to an individual C3 system or require a comparative evaluation of several alternative systems and/or forces?

6. What threat and scenarios are appropriate and available?

7. What part of the life cycle of the C3 system is involved? Time frame?

8. What level (system, subsystem, platform, force, etc.) is the analysis focused upon?

9. What type of measure, i.e., how quantitative, will answer the decision maker's question?

10. What analytical support will be required? Testing? Simulation?

61

In summary, three steps take place in Module 1: (1) the decision maker's needs are characterized; (2) the problem's scope and depth are selected; and (3) the remaining modules are previewed for their potential impact on the problem statement and analytical effort required.

2.   Module 2: C3 System Bounding

Module 2, as described by Figure 3, enumerates the relevant system elements that bound the problem of interest. The first goal is to delineate the difference between the system being analyzed and its environment. To bound the C3 system, the analyst should employ the three-part definition, based upon JCS Publication 1. In it, a C3 system consists of: (1) physical entities—equipment, software, people and their associated facilities; (2) structure—organization, concepts of operation, standard operating procedures, and patterns of information flow; and (3) process—the functionality or "what the system is doing" which is pursued in Step 3. In the second module the C3 system, identified by its human, hardware and software entities and structures, is related to the forces it controls and the environmental stimuli to which it responds, including the enemy. Once the system elements of the problem have been identified, the C3 system of interest may be further bounded by relating the "physical entities" and the structure components to the graphic representation of the levels of analysis, using the graphic model as shown in Figure 4.
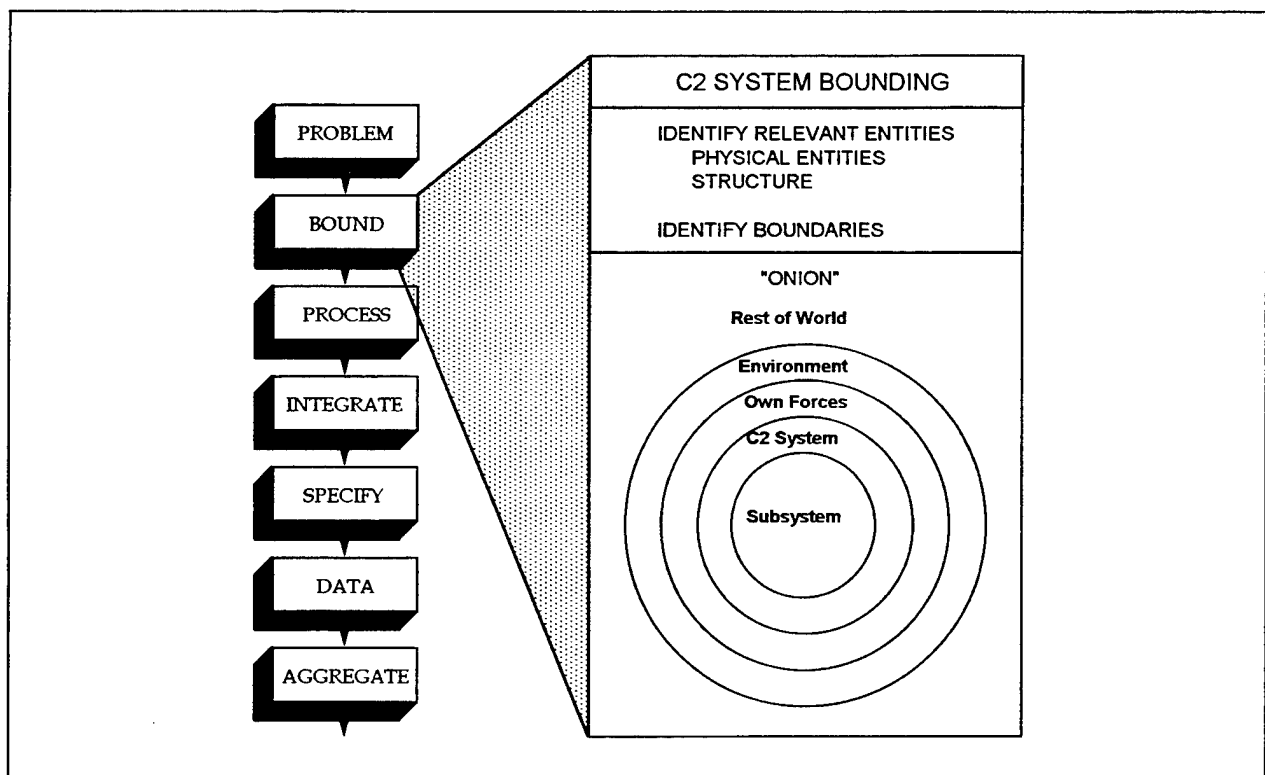
Figure 3.   MCES C2 Systems Bounding

This series of levels is referred to as the "onion skin." In the most inclusive depiction of this graphic, there are five rings. Beyond the outer ring is the rest of the world, which essentially relates to elements and structure that exist and may have import with respect to similar problems, but which are outside the scope of the problem at hand. In contrast, the outer ring represents the environmental factors that require explicit assumptions in the problem. This ring may be seen as

including the major scenario components. The next ring, moving inward, deals with the forces under influence of the C3 system upon which the evaluation is centered. The C3 system itself is the focus of the next ring, and its component subsystems make up the innermost ring. As is clear from the foregoing, this graphic is a structured static display of the physical entities.
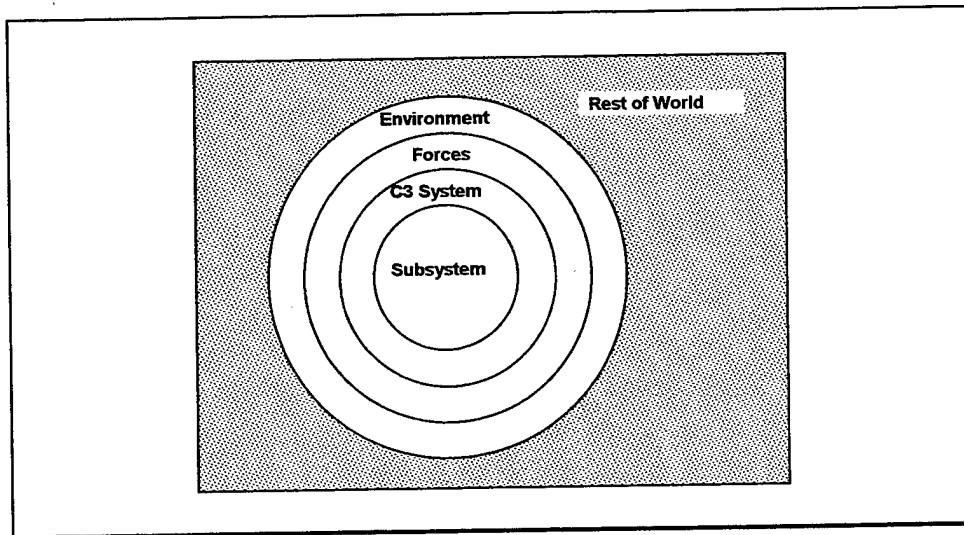
Figure 4. C3 System Bounding and Level of Analysis

In summary, 1) the C3 system statics must be distinguished from the C3 system dynamics, the "C3 process" and its functions. 2) The statics must be be listed as the physical entities together with the structural relationships of C3. 3) The structure is represented by the customary physical arrangement and interrelationships of entities in the form of command structure, the standard operating procedures, protocols, message formats and reporting requirements. Bounding the C3 system often leads to broadening the system of interest. It may be necessary to consider the source of information as well as the display that is being decided upon in a particular decision.

3.   Module 3: C3 Process Definition

After the system is bounded and the system elements identified, the dynamic C3 processes of the system are identified as noted in Figure 5.
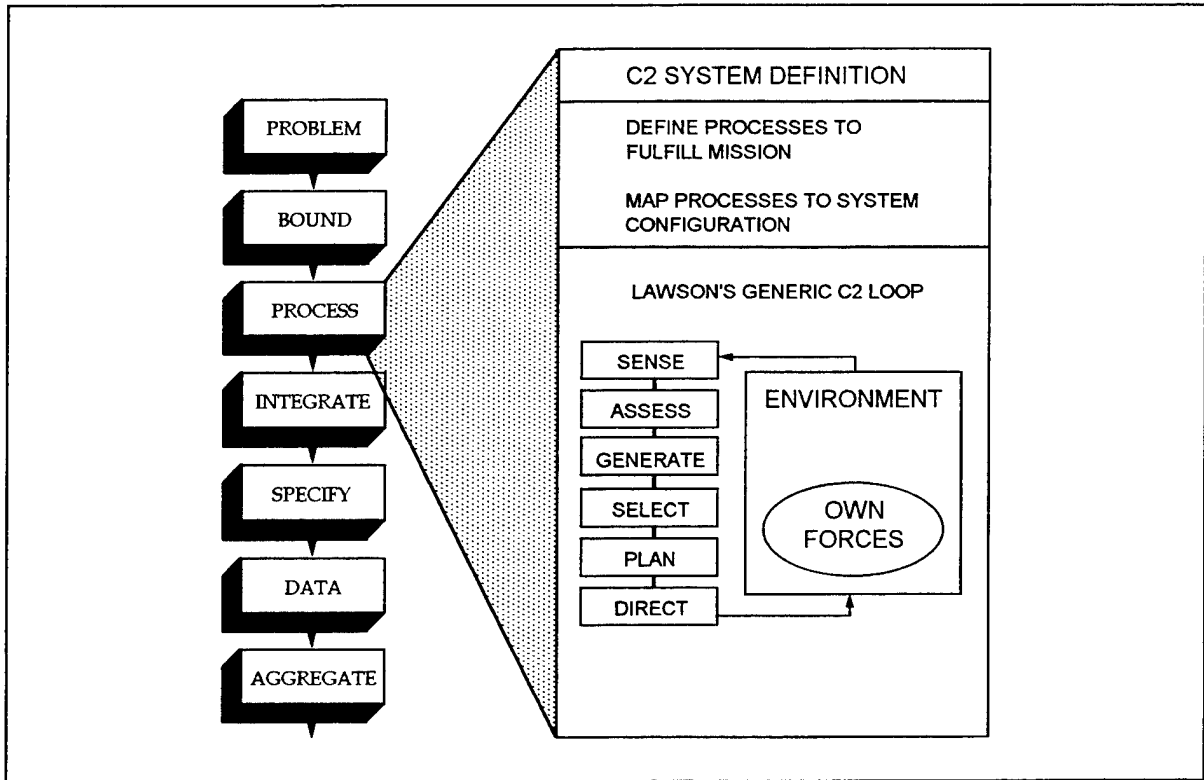
Figure 5. C2 Process Definition

Module 3 focuses the analyst's attention on: (1) the environmental "initiator" of the C3 process, which results from changes in the desired state, usually of enemy forces; (2) the internal C3 process functions (sense, assess, generate, select, plan, direct); and (3) the input to and output from the internal C3 process and the environment. The C3 process functions are generic and may be adapted to the specific functions of air defense, ground operations etc. They can be described briefly here as six function.

- Sense—the function that collects data necessary to describe and forecast the environment, which includes:

  (1) The enemy forces, disposition and actions.

  (2) The friendly forces, disposition and actions.

  (3) Those aspects of the environment that are common to both forces—for example, weather, terrain and neutrals.

- Assess—the function that transforms data from the sense function into information about intentions and capabilities of enemy forces and about capabilities of friendly forces to determine if deviation from the desired state warrants further action.

- Generate—the function that develops alternative courses of action to correct deviations from the desired state.

- Select—the function that selects a preferred alternative from among the available options. It includes evaluation of each option in terms of criteria necessary to achieve the desired state.

- Plan—the function that develops implementation details necessary to execute the selected course of action.

- Direct—the function that distributes decisions to the forces charged with execution of the decision.

In summary, these six functions have been found to be sufficiently comprehensive to map to almost any C3 process. They are applied iteratively.

4.   Module 4: Integration of System Elements and Functions

As noted in Figure 6, in Module 4 the relationships between the physical entities and structures (defined in Module 2) and the C3 processes or functions (described in Module 3) are first identified and described—who does what, when. Then techniques such as PERT charts, data flow diagrams or
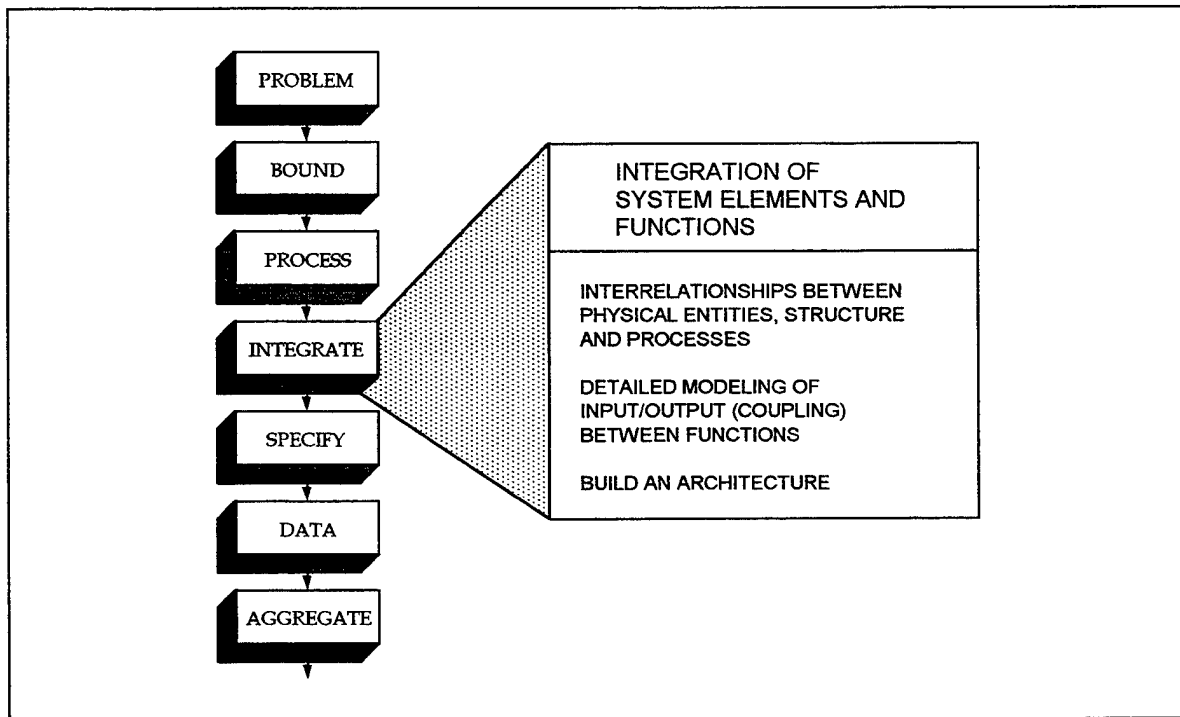
Figure 6. Integration of System Elements and Functions

Petri nets may be used to model the messages or information flows that are used to control these relationships. Information flows support decisions that link the separate C3 functions into the architecture containing the relevant C3 system. The term "architecture" is used to describe the output of module 4 to emphasize the integration via defined interfaces and standards of the individual C3 subsystems. The physical entities, structures and functions of these individual systems are coherently controlled in a dynamic architecture. The architecture might indeed become a functioning computer model of the system which would support an evaluation of mission effectiveness. The final form of the architecture will at least include the process description of the system elements performing the processes arranged in a structural framework as indicated in Figures 3-4. These may be adequate to support qualitative evaluation of the

65

architecture. A quantitative description of the elements and the inputs to the processes are required even if a model cannot be built in the time available. Even these descriptive inputs allow an informal assessment on a subjective basis. In summary this module maps Steps 2 and 3 together and provides quantitative information preferably as a model of the architecture in a static and/or dynamic mode.

5.   Module 5: Specification of Measures

A C3 measure can usually be categorized as either a performance measure or a vulnerability measure. There are generic sets of both of these categories such as the TRI-TAC MOEs shown in Table 1. These TRI-TAC measures are generic and need additional specification in terms of a particular scenario and C3 system. For example, the units of speed of service, interoperability and survivability must be identified with reference to the mission and level of the system.

TABLE 1. TRI-TAC MEASURES OF EFFECTIVENESS

PERFORMANCE MEASURES	Grade of service Information Quality Speed of Service Call Placement Time Service Features Lost message Rate Spectrum Utilization Transportability Mobility Ease of Reconfiguration Ease of Transition Interoperability
VULNERABILITY MEASURES	Index of Survivability (Overt) Index of Survivability (Jamming) Index of Availability Interrupt Rate Security

In Module 5, as illustrated in Figure 7, the analyst specifies the measures necessary to answer the problem of interest as defined in Module 1 and in the system bounding process and integration. The component levels and functions of the C3 system definition modules may be examined to derive an initial set of relevant measures, which are then subjected to further scrutiny: (1) comparison with a set of criteria, Table 2, which may reduce the number to a more manageable set; (2) the remaining measures are then classified as to their level of measurement (MOFE, MOE, MOP or parameter) which may lead to association of some to a lower level than currently of interest; (3) mapping of the MOFE to related MOEs and then to related MOPs, etc., and (4) the resulting high level measures are examined for the practicability of measuring alternative configurations of the physical entities, structure and/or processes of the C3 system in the scenarios defined in Module 1. Practicality often drives measurement down to the level of MOE or even MOP because combat oriented measurements are inherently difficult.

66

TABLE 2. CRITERIA FOR EVALUATION MEASURES

CHARACTERISTICS	DEFINITION
Mission-oriented	Relates to force/system mission
Discriminatory	Identified real differences between alternatives
Measurable	Can be computed or estimated
Quantitative	Can be assigned numbers or ranked
Realistic	Relates realistically to the C2 system and associated uncertainties
Objective	Can be defined or derived, independent of subjective opinion
Appropriate	Relates to acceptable standards and analysis objectives
Sensitive	Reflects changes in system variables
Inclusive	Reflects those standards required by the analysis objectives
Independent	Is mutually exclusive with respect to other measures
Simple	Is easily understood by the user

Each of the three levels of the C3 system in the onion-skin diagram is directly related to measures of performance (MOPs), measures of effectiveness (MOEs), and measures of force effectiveness (MOFEs) as shown in Figure 7.
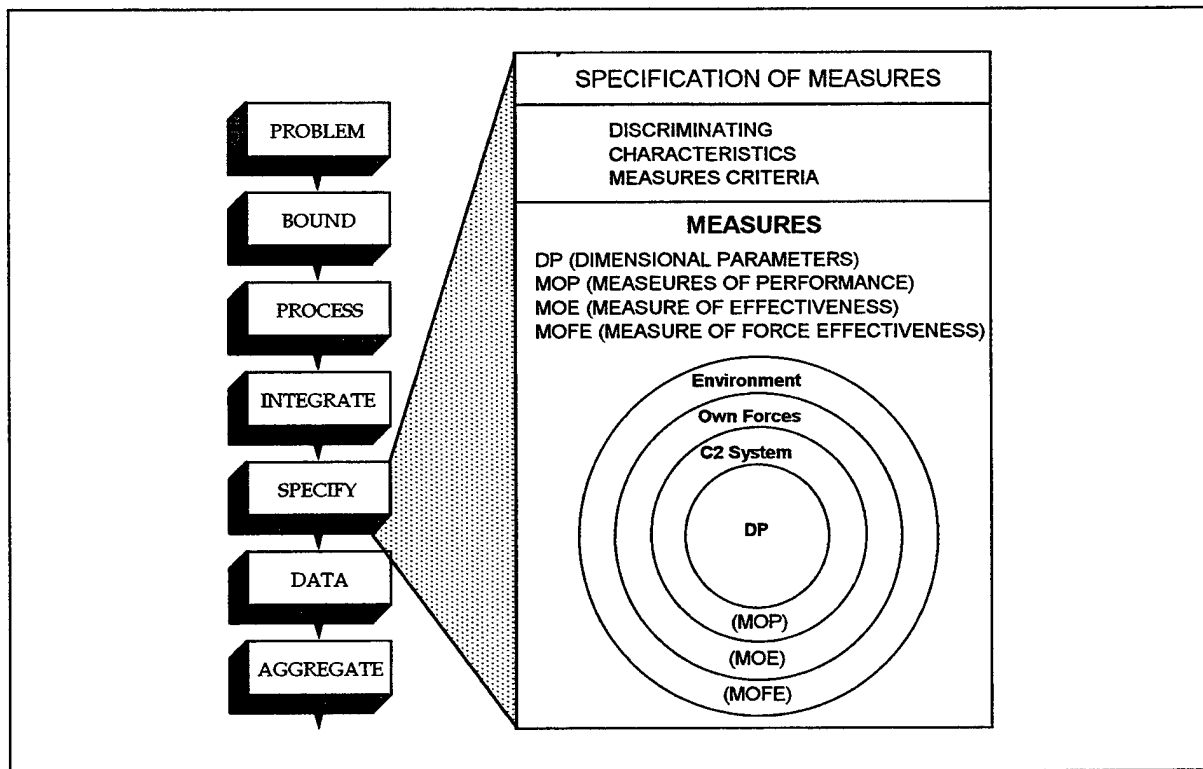
Figure 7. Specification of Measures

The determination of the boundary helps to identify what level of measure is appropriate. If the boundary between the force and the environment is of interest, measures of force effectiveness (MOFE) are required. Dealing with the boundary between force and the C3 system leads to measuring the effectiveness (MOE) of the C3 system. At the subsystem level—that is

67

within the boundary of the system—are measures of performance (MOP) of the functions. Finally, within the subsystem are Dimensional Parameters (DP). Measures at the higher level, MOFEs and MOEs, are most desirable because they are closer to the ultimate purpose of the C3 system and because they summarize many of the lower level measures in a meaningful way.

In summary, this module's implementation results in the specification of a set of measures that is focused on the C3 process functions within the C3 system, the overall performance of the C3 system and on the force effectiveness of the C3 system combined with the forces and weapon systems, if at all practical.

6.   Module 6: Data Generation

The generation of values for the measures determined in the previous module is addressed by the sixth module. These values are the result of the implementation of this module as noted in Figure 8. Here, one of several types of data generators such as exercises, experiments, simulations, models or subjective judgement is selected. The MCES accommodates a variety of data generators. The prime requirements are that the data generator is: (1) available to the analysis; (2) focused on the mission area/analysis objectives of the evaluation; and (3) adaptable to produce, with minimal modification, the values associated with the measures specified in the previous module. The analyst must consider the following: reproducibility of results, precision and accuracy, costs and timing of data collection, environmental controls, and experimental design in the final choice of how to generate the values.
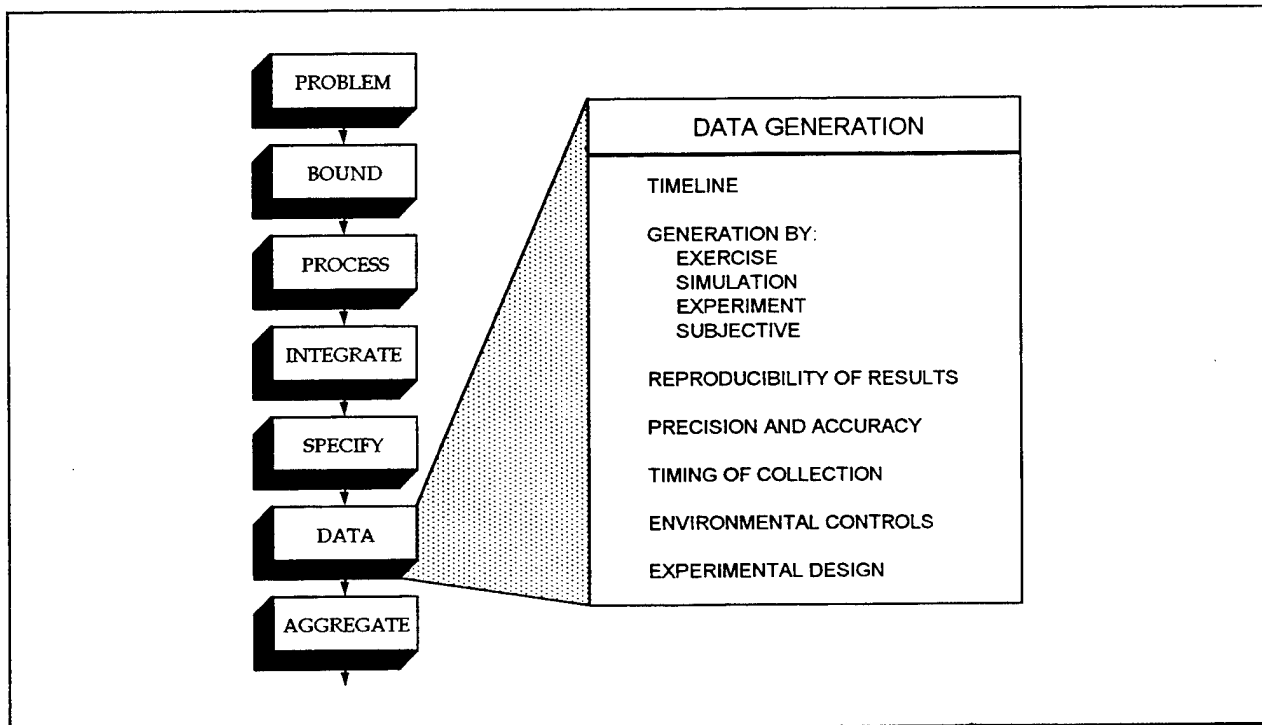
Figure 8. Data Generation

This step is directly supported by Module 4, the integration of elements and processes. If the integration has resulted in a quantitative model it will be straightforward to generate output data. The verification of input data from modules 2 and 3 and validation of the model must also

68

be addressed. Alternatively, if only a conceptual mapping of function to structure is accomplished in Module 4, the generation of values for measures may be only a qualitative comparison table or relative judgmental statements by experienced personnel.

In the typical implementation, the relationships established in module 4 are translated into computer code. In this process it will often be necessary to define additional relationships and obtain more input data. The validation and verification of this code as a representation of the problem must also be addressed. The National Test Bed's Confidence Assessment Methodology is a recommended reference for this step.

7. Module 7: Aggregation of Measures

In Module 6, Data Generation, the analyst obtains values for the specified measures which will be analyzed and interpreted in this module as noted in Figure 9. Because varying scenarios may be important for each iteration of the MCES, the analyst must determine the importance of each factor. The final module addresses the issue of how to aggregate and interpret the measures. Three levels of measurement (performance, effectiveness and force effectiveness) with multiple values from each level may be available. The current state of the art requires that both qualitative (such as red-yellow-green charts) and quantitative (such as utility weighting) aggregation techniques be considered.

The nature of the problem and available tools determine the mix of these techniques. Different problem areas addressing different decision makers' analytic needs will result in differing requirements for aggregation of constituent measures, but the mappings between levels allow the decision maker to make an informed decision and understand the reasons for it. The issues of measure causality, sufficiency and independence must be considered. The analyst must decide if the decision maker's original queries have been addressed by the MCES analysis. Finally, suitable graphics should be prepared for interaction with the decision maker.
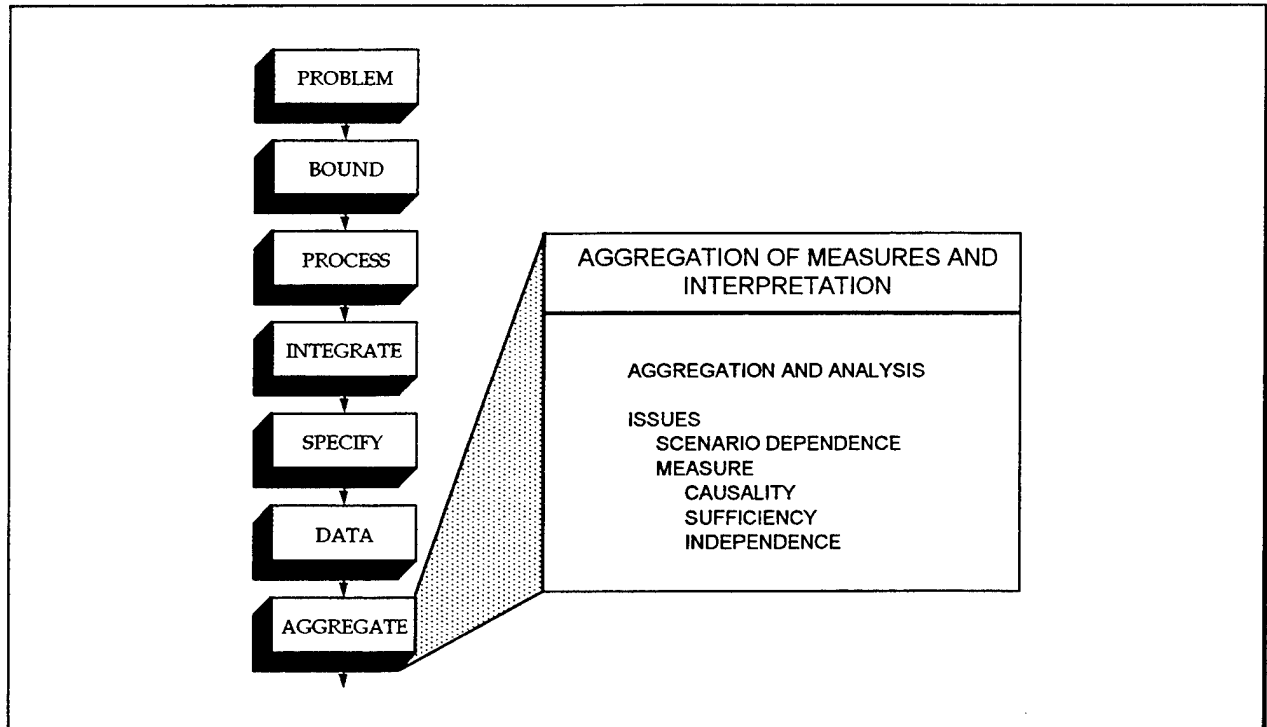
Figure 9. Aggregation and Interpretation of Measures

The implementation of this module provides the analytical results tailored to address the problem posed at the beginning of the procedure. The results, made up of the aggregated values and measures, should be provided to the decision maker in a format that will expedite his consideration of the analysis. Whenever appropriate, graphics are used to summarize and show trade-offs.

Finally the results are provided to the decision maker. Two courses of action are available. First, the decision makers may identify the need for further iteration. Or they may proceed to implement the decision. In most situations, explanation of objectives and the reasoning behind the decision help the implementation by lower levels of the organization. MCES is an aid in conveying the context, structure and evidence supporting the decision to these levels.

A2. Navy Major-Caliber Ammunition Reliability Goals

1990-1997
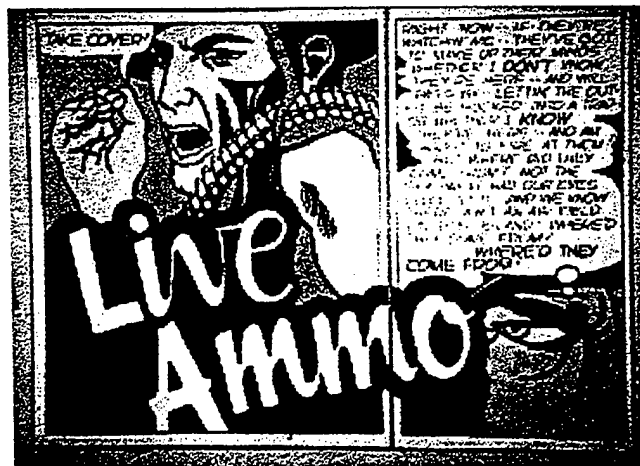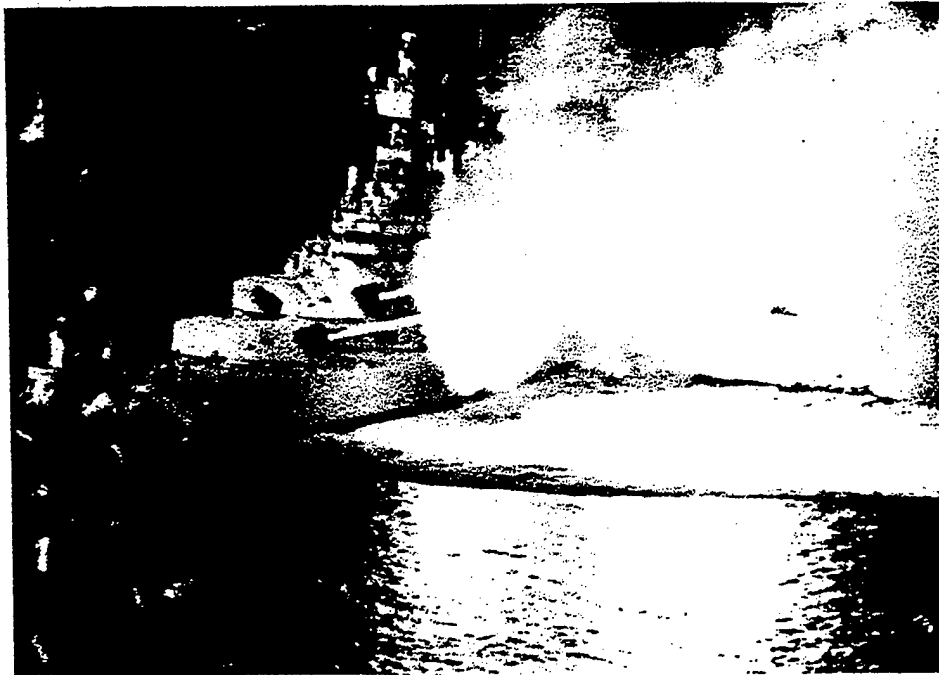
by

Michael P. Bailey, Ph.D.
Alexander J. Callahan, Jr., M.S., P.E.
Ammunition Working Group
Department of Operations Research
Naval Postgraduate School

ACKNOWLEDGEMENT

Foreword

The United States Navy purchases millions of dollars worth of major-caliber ammunition each year to supply its warships. Combined, U. S. services buy hundreds of millions of dollars of ammunition annually. The rounds of ammunition are purchased by component, and failure of each component has a different effect on the gun system. Thus, establishing component reliability thresholds is a complex and important task. We describe the decision process for establishing the threshold reliability for components of navy major-caliber ammunition. We present a measure of reliability performance, called ef^*, which relates directly to the weapons system's performance in a naval gunfire support environment. We use a simulation model to establish this relationship, a regression metamodel to estimate its parameters, and a simple decision process to specify component reliability thresholds which ensure that the ammunition is mission effective. We provide a summary of the data collected between 1990 and 1996. We report the results of analysis and trends of the Marine Corps Scenario 6 using the 1991 to 1993 data and then the 1994 to 1996 data. We discuss the need for continued data collection, analysis, and the requirements to upgrade assessment methodology to conform with High Level Architecture guidance and the Major Regional Conflict West Scenario.

Table of Contents

Table of Figures

1. INTRODUCTION

In 1990, the Naval Postgraduate School (NPS) Ammunition Working Group, at the request of the Naval Surface Warfare Center (NSWC) Crane Division, began development work on a mathematical model to find the relationship between effectiveness (ef) values and battle goals. The goal of the original work was to determine minimum reliability thresholds for the components of major caliber gun ammunition with the anticipation that the results would impact program management decisions regarding procurement and improvement of gun systems components.

A common belief was that the reliability of each of the round components uniquely impacts the effectiveness of the weapon system, and subsequently the effectiveness of the battle force. To procure and maintain ammunition that will provide adequate utility to naval forces, there must be a clear understanding of the relationship between ammunition component reliability and force effectiveness. It was further believed that this relationship should guide decisions in procurement and surveillance.

Aside from the immediate ramifications of a failed-round component, that being that the round is ineffective, there may also be significant negative effects in the continued operation of the gun system. For instance, if a propellant component fails, the projectile remains stuck in the gun chamber. If the gun has been firing continuously for some time, the gun chamber may be quite hot, thus the explosive charge within the stuck projectile becomes a safety hazard. Its removal from the gun is a delicate operation that takes a significant amount of time, and causes the ship's execution of its mission to be delayed. Other types of round component failures cause different negative effects on the ship's performance. Thus, any effort to establish ammunition component reliability must take into account the complex relationship between the impact of each failure type on operational performance. This impact can be measured in terms of system delay.

In this seminal work, a mathematical model led to the computer simulation used to determine reliability thresholds for major caliber gun ammunition. The simulation models the action of the 5 inch 54 Mark 45 (MK45) gun systems of the battle force, the cycle times associated with the crew performing NGFS tasks, the variability of miss distance, the variability of navigation error, and the variability of spotting. The simulation uses replications of the gun system to represent the ships specified by the study plan. The actions of the Naval Gun Fire Support Group are then simulated against the prescribed targets The output shows the relationship between ef and the scenario time.

A decision process was designed that used the simulation to relate the effects of the reliability of ammunition components to the performance of the gun weapon system in particular, an expeditionary battle force in general and

prescribes component reliability thresholds required for that force to meet the threat prescribed by policy and doctrine.

2. METHODOLOGY AND MODEL APPROACH

We provide a methodology for determining the minimum reliability thresholds for components of major-caliber ammunition. We do this by providing the decision maker with a functional relationship between component reliability and mission performance. By using existing mission performance thresholds, the decision maker can prescribe minimum component reliability goals. Mission performance is measured in two ways for a naval gunfire support (NGFS) mission.

1. mission time ---- the time required to destroy all of the targets on the scenario target list;

2. average casualty rate ---- the rate of casualties inflicted by the opposition during the amphibious assault;

We modeled the NGFS mission because it employs gun systems in sustained operation where the impact of system failure is most acute. The two measures of performance above are referred to as *battle goals*. The battle goals, also called standard performance measures, are those goals specified as indicators of a ship's ability to successfully complete an assigned war plan. Mission time or Scenario time, the length of time required to neutralize the standard set of targets in the specified Amphibious Operations Area (AOA), was used to evaluate the reliability goals. Mission time is essential to test the battle force mission area capability for Naval Gunfire Support (NGFS) because success of the theater operations depend on a successful Amphibious operation in a restricted time frame.

The decision maker is willing to specify the acceptable values for the battle goals. Hence, our objective is attained if we can translate performance with respect to the battle goals into reliability thresholds. The approach we took is illustrated in Figure 1. This structure is similar to some of the hybrid analytic/ simulation models with general structures described in Sargent and Shanthikumar [9]. What makes this process interesting is that there are an infinite number of mixtures of ammunition component reliability which produce identical battle goal values. What we strive for is a method of summarizing all of the ammunition component reliability data into a single number which represents the degree of impact that a particular failure has on the battle goal values as well as the probability of the failure occurring.

Let us trace through the process shown in Figure 1 for determining reliability thresholds for a particular scenario. Starting at the top, we form a list of the gun types, the types of ammunition, and the targets used in the scenario. From here, we collect data concerning the failure mode and accuracy behavior of the guns in the fleet. We construct a planned movement pattern for the ships as they engage the targets. Last, we choose values for the reliability of each; component of each type of ammunition round used in the scenario. All of these data, including the round reliability, guns, motion, accuracy, and targets, form a set of computer files we call the NGFS scenario. This scenario is input into our NGFS simulation model, which produces observations of the battle goal values. This is described in detail in Section 4.
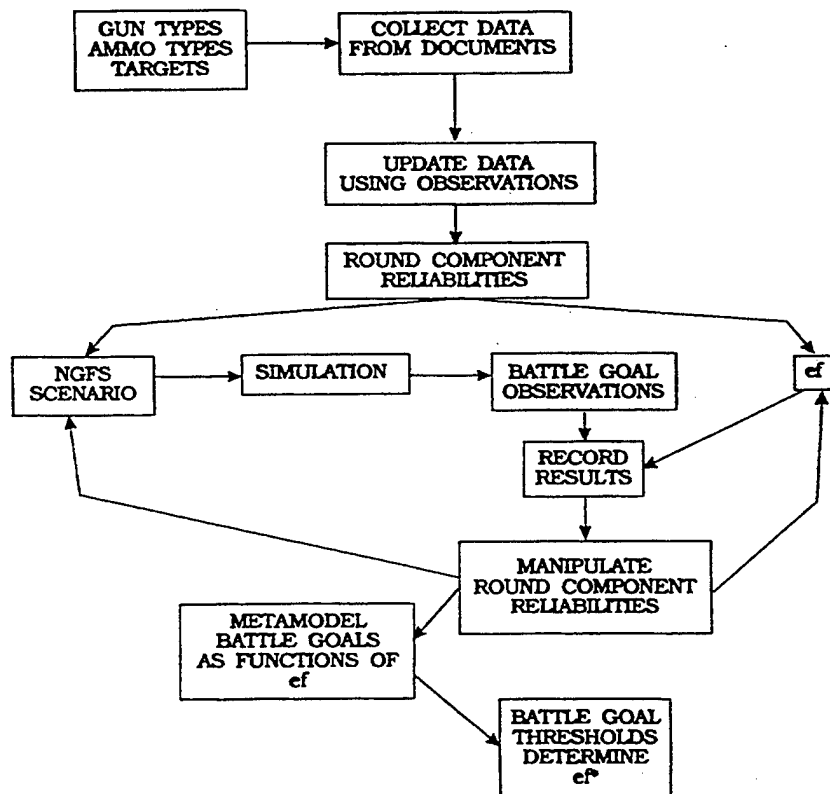
Figure 1   Flow diagram of reliability goal determination

Simultaneously, the value of the static measure *ef is* calculated from the gun and ammunition reliability values and the system's recovery times. After replicating the engagement several times and recording the outcome of the battle goal observations for a single value of *ef,* we then manipulate the reliability of one or more of the ammunition components and produce another set of battle goal observations and another value of *ef.* This measure and its calculation are presented in Section 3.

Repeating this process several times, we produce the data to support the construction of a metamodel of the battle goal outcomes as a function of *ef.* Finally, we use the metamodel to translate acceptable battle goal values into a threshold value for *ef,* which we call *ef*.* In Sections 5 and 6, the metamodel and its application are described. As we shall see in the next section, *ef is* very simple to calculate, and depends only on reliability and recovery time data --- no scenario is required. We hypothesize that *ef* explains most of the impact that round component reliability have on the system's behavior with respect to battle goals. Thus, we have a means of comparing different component reliability configurations, as well as specifying component reliability thresholds. Furthermore, the functional form of *ef* combined with the functional form of the metamodel constructed lead us to some interesting conclusions about equipment reliability and training levels for naval gun system repairmen.

3. MEASURING PERFORMANCE WITH *EF*

In this section we establish an appropriate performance measure to predict variations in the effectiveness of the NGFS system with respect to changes in round component reliability. This measure, called *ef, is* used in the metamodel as a predictor of battle goal performance.

From a system reliability point of view, a round of ammunition is a very simple device. It is a series system with a few fuses and charges, along with some sort of sensor. The ammunition comes in two pieces: the projectile containing a fuse, the sensor, and an explosive charge; and a propelling charge, which launches the projectile out of the barrel of the gun. A simple diagram is shown in Figure 2. Each of the components of the round must operate successfully for the round to operate. Figure 3 shows the flow of energy through those components of interest.

Serial operation causes the observed failure rate of a component in the fleet to be different than the defect rate of the component when purchased. To see this, suppose that the fuse charge of a round fails to operate. In this case, the burster charge of the round is not observed *whether it is defective or not.* Hence, fuse charge failures *mask* burster charge failures. This same phenomenon causes gun failures to mask round failures.

Let us classify the failures experienced by the gun/round system into N categories such that a failure of type T_i causes the gun to stop firing for some deterministic time T_n. Let T be measured in time units equal to the time required to fire the gun once. The nature of this system allows us to number components so that the lower-numbered components take precedence over the higher levels.

The defect rates that can be estimated by destructive testing of individual components are given by

$x_1 = P[\text{failure 1 occurs}]$,

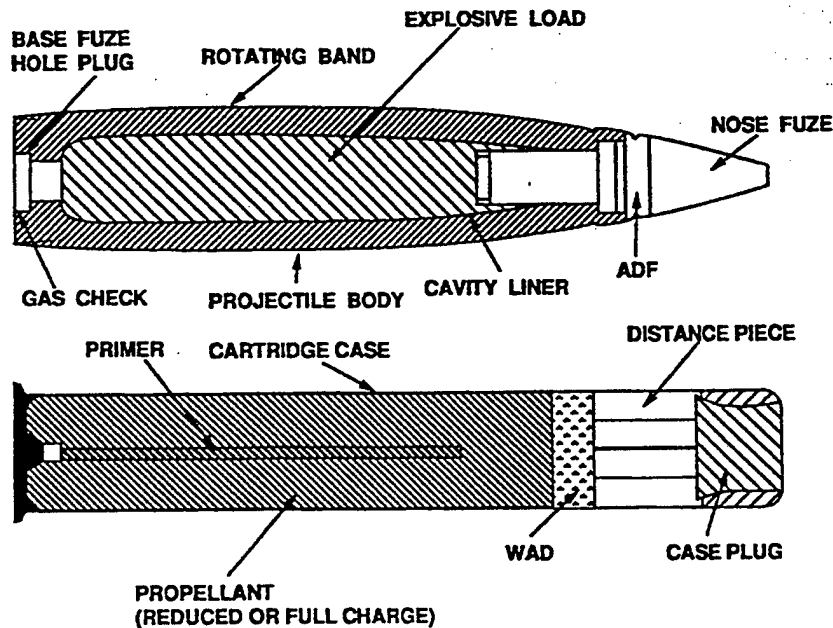$x_i = P[\text{failure i occurs failure j does not occur}, j < i]$,

Figure 2   Five-inch, 54-caliber propelling charge and projectile

for $i = 2, \ldots, N$. The values of the x_i's come from acceptance tests or from destructive testing of the round components. The probability that a failure is observed in the field is given as

$q_i = P[\text{failure i occurs and failure j does not occur for } j < i]$.

Thus,

$$q_i = X_i \prod_{j=1}^{i-1} (1 - Xj), \tag{1}$$

for $i = 1, 2, \ldots, N$. Note that the set of events associated with $q_1, q_2, \ldots q_N$ are mutually exclusive, so that $1 - \sum_{i=1}^{N} q_i$ is the probability that the gun/ round system experiences no failures when fired.

Let F_i be the number of failures of type i occurring between successful firings of operational rounds. Let $p_i = 1 - q_i$, $i = 1, 2, \ldots, N$. and $p = [p_i, P_2, \ldots, P_N]$. Let

79

the random variable T(p) represent the time required to fire a single effective round.

Then,

$$T(p) = \sum_{i=1}^{n} [F_i T_i] + 1,$$ (2)

where 1 is the time required (in the time units used to measure T_i) to fire a round from the gun with no failure.

Figure 3   The round as a simple series system.

Because each round is assumed independent of all others, F_i is a geometric random variable with parameter P_i, i = 1, 2, . . ., N. and the F_i are mutually independent. Thus, the first two moments of T(p) are given by

$$E[T(p)] = \sum_{i=1}^{N} (T_i q_i / p_i) + 1,$$ (3)

$$VAR[T(p)] = \sum_{i=1}^{N} (T_i^2 q_i / P_i^2)$$ (4)

In comparing values of $E[T(p)]$ for two different round designs, the round design with the smaller $E[T(p)]$ seems better to most reasonable analysts. However, it is possible to construct an example where a round has smaller $E[T(p)]$ but larger VAR[$T(p)$]. We discuss the ramifications of this phenomenon in a later section.

We now define our measure of performance of the gun-round system as

$$ef = 1 / E[\,T(p)].$$ (5)

This quantity gives the expected number of successfully fired operational rounds per unit time. As we shall demonstrate, simulation results confirm that *ef* represents a good measure of performance of the NGFS system, capable of accurately predicting our battle goal performances.

The functional form of *ef* provides some interesting insights immediately. We can explore the effects of small changes in the failure probabilities or recovery times by taking partial derivatives:

$$\delta ef / \delta p_i = (\,T_i / p_i^2\,)\, ef^2,$$ (6)

$$\delta ef / \delta T_i = ((p_i - 1) / p_i) \, ef^2 \qquad (7)$$

The improvements in system reliability realized by incremental improvements of component reliability are ordered by the ratio T_i / p_i^2. If T_i is large and x_i is large (making P_i small), then this ratio becomes large. Stating that often-occurring, long-recovery-time failures are important coincides with our intuition, but precisely ordering the components in importance is facilitated by the development of *ef*.

Furthermore, since $0 \leq P_i \leq 1$ and $ef \geq 0$ we know that *If $\delta ef / \delta T_i \leq 0$*. This relationship concerns the training of the gun's crew and the design of the failure recovery system. We see that the greatest operational payoff per unit of reduced recovery time comes from the failure with the lowest P_i. Thus, failures which occur most often are most critical, independent of the length of the recovery. These two conclusions will be used by our sponsors in making decisions regarding design of rounds, training of crews, and granting of bonuses to manufacturers who produce exceptionally reliable components.

Further exploring the relationship between training, component reliability, and operational performance, consider a component j which we can either make more reliable (i.e., decrease x_j) or make it repairable in less time (i.e., decrease T_j). Equating *ef* for two systems, one with better reliability for component j and one with shorter repair time for component j, we have

$$ef_1 = \left[\sum_{i=1, i \neq j}^{N} \frac{q_i}{p_i} T_i + \frac{q_j}{p_j} (T_j + \partial) \right]^{-1}, \qquad (8)$$

$$ef_2 = \left[\sum_{i=1, i \neq j}^{N} \frac{q_i}{p_i} T_i + \frac{q_j - \epsilon}{p_j + \epsilon} T_j \right]^{-1}, \qquad (9)$$

$$p_j + \epsilon = 1 - (x_j + \gamma) \prod_{i=1}^{j-1} (1 - x_i), \qquad (10)$$

where δ is the (negative) increase in repair time, ϵ is the increase in p_j, and is the (negative) increase in the defect rate of component j. Solving, we get

81

$$\delta = \frac{-\epsilon T_j}{(p_j + \epsilon)q_j} \tag{11}$$

$$= \frac{T_j \gamma \Pi_{i=1}^{j-1}(1 - x_i)}{q_j(p_j - \gamma \Pi_{i=1}^{j-1}(1 - x_i))}. \tag{12}$$

table 1 presents the results of a simple numerical experiment. We start with a baseline system with a component j which has a 1% defect rate, and we determine the equivalent repair time reduction for a proposed component which has either a 0.1% or 0.8% defect rate, and we do this for components which have baseline repair times of 10 and 100 time units. Thus, the top entry tells us that if the component is improved so that its defect rate is 0. 1%, we are providing as much improvement to the system as if we were to reduce the repair time from 10.0 time units to about 1.0 time units. Notice that the impact of the defect rates of components 1, 2, . . ., j - 1 are minimal here, and that the repair time reduction is a constant proportion of the original repair time.

Table 1 Tradeoff between improved repair time and improved defect rate.

T_j	$1 - x_j$	$\Pi_{i=1}^{j-1} (1 - x_i)$	p_j	γ	ϵ	δ
10.0	0.99	0.95	0.9405	-0.009	0.00855	-9.01
				-0.002	0.0019	-2.02
		0.8	0.992	-0.009	0.0072	-9.01
				-0.002	0.0016	-2.01
100.0	0.99	0.95	0.9405	-0.009	0.00855	-90.1
				-0.002	0.0019	-20.2

4. THE SIMULATION MODEL

Our sponsor, the Naval Surface Warfare Center, Crane Division, Crane, Indiana, provided us with an NGFS scenario that included fleet composition, target types and locations, and deadlines. The simulation we developed inputs this scenario, then subjects the targets to realistic prosecution.

We simulated the spotting procedures used by naval gunners, the allocation of targets to ships and guns, the navigational and guidance errors of the correct magnitude and probability distribution, the random effects of the rounds' impact on the targets, and failures of ammunition, guns, and crew. We modeled realistic delays for battle damage assessment, spotting calculations, communication transmission, and command decisions. This system was modeled in the MODSIM [8] object-oriented simulation programming language.

Data supporting the simulation was of two types, engineering data and tactical data. Engineering data includes the frequency of each failure mode and the associated recovery time. This data was drawn from the U. S. Navy's Material Readiness Database (MRDB) [7]. The logic used in simulating the repair process came from Weapons System Fundamentals [10]. We also require damage data for each pairing of ammunition type with target type from the joint munitions effectiveness manuals (JMEMs) [4]. The damage model we use takes

· range from the ship to the target

· shell type, including fuze type

· target type

· range from the point of shell impact to the target's center

and produces a fraction of expected damage done to the target. As seen in Figure 1, all of this data is submitted to the simulation model. A subset of this data, the failure probabilities and repair times, are also used to produce ef using (2) and (5).

Required tactical data include all of the information about how the ships in the task force behave when prosecuting the targets. The *Gunsmoke Manual* [2] provides reasonable ship maneuvering regimes for the target engagements in the scenario. Other required data about the engagement was supplied by unscientifically surveying numerous surface warfare specialists at the Naval Postgraduate School.

For the purposes of exposition here, we developed a hypothetical scenario involving a small task force of five (5) ships executing preassault engagements in a harbor area with twenty-four (24) targets. These targets included an airfield, several reinforced command and control bunkers, artillery positions, and troop positions on the beach. The preassault phase is supposed to take one full day (1440 minutes). The mission includes several counterbattery actions, where a ship disengages from its current target and engages a target shooting at the ship. The model and data were validated by comparing model outcomes for single target engagements with qualification data from the Atlantic Fleet Weapons Training Facility (AFWTF), Isle de Vieques, Puerto Rico. This comparison was done based on mission times for these single-target engagements.

From each replication of the simulation, we collected the values of the battle goals described in Section 2. Let there be W targets in the scenario examined, and suppose that $f_i(t)$ is the firepower of target i which survives at time t, $i = 1, 2, \ldots, W$. The firepower of a target is measured in expected casualties per time

unit. If a target is not a direct combatant, a radar site for example, its firepower is estimated as the difference in expected casualties with and without the system. Let

$$f(t) = \sum_{i=1}^{w} f_i(t) \qquad (13)$$

be the total casualty-inflicting capability of the enemy at time t. Our measures of performance are

$bg^{(1)}$ = the time that all of the targets in the scenario are destroyed $\qquad (14)$

$= \min t : f(t) = 0; \qquad (15)$

$bg^{(2)}$ = the time-integrated firepower of the surviving targets $\qquad (16)$

$$= \int_{assault\ time}^{bg^{(1)}} f(t)\ dt. \qquad (17)$$

The assault time used in the calculation of $bg^{(2)}$ is the time that the troops the NGFS is supporting are scheduled to arrive in the target area. This time is given in each scenario, and is not allowed to vary with the sample path. That is, the troops will arrive on the beach at the required time and are subjected to the remaining enemy firepower until it is extinguished.

5. REGRESSION METAMODELING

The final steps in the decision process shown in Figure 1 are to estimate the functional relationship between ef and the battle goals; and to use thresholds for the battle goals to determine the minimum acceptable ef.

After an initial plot of the $BG^{(i)}$ vs. $1/ef$, we concluded that a simple linear relationship between $BG^{(i)}$ and $1/ef$ was plausible. Regression metamodeling is the practice of fitting a functional model to the output of a computer simulation, see [5]. Figure 4 shows this plot for the time-integrated firepower battle goal $BG^{(2)}$. In the case of each battle goal, the regression passed all the usual tests for model fit, normality of residuals, and significance of the slope estimate. Some points suspected of being leverage points were investigated further and found

to not influence the regression estimates too much. However, there is an observable amount of variation in the responses not explained by ef. The sources of this variation are the random effects not encapsulated in ef, namely, navigation and guidance errors, focusing and diffusing firepower on the targets, effectiveness of rounds against targets, and delays caused by calculation,

communication, and command. For this example, we estimate the pair of regression equations

$$bg^{(1)} = 45.79 + 41.20 \; \frac{1}{ef}, \qquad (18)$$

$$bg^{(2)} = 235.36 + 199.95 \; \frac{1}{ef}, \qquad (19)$$

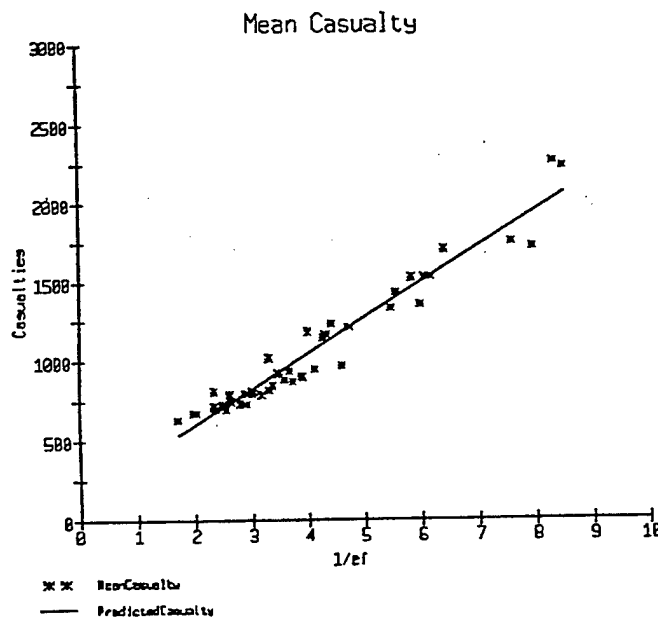Figure 5 shows plots of observed battle goals with *ef*.

Figure 4 BG$^{(2)}$ regressed against the value of 1/ef

We now have established the functional relationship between defect rates in ammunition components and battlefield performance. In the next section, we will complete the development of our reliability goal determination process.

Suppose the decision maker is willing and authorized to establish that the mission described in the scenario should take no longer than 150 minutes, and

that the time-integrated firepower is not acceptable if it is above 860. The reader should note that this setting is hypothetical.

Using the regression result in (18) and (19), we see that the 150-minute deadline requires *ef* to be above 0.40, while the time-integrated firepower of 860 forces *ef* above 0.32; *ef** = max[0.40, 0.32] = 0.40. In this case, the mission time constraint clearly restricted the value of *ef* more than the casualty constraint. In cases where both measures constrain *ef* to nearly the same degree, the analyst must be cognizant of the dependence of $BG^{(1)}$ and $BG^{(2)}$. A more sophisticated linear model which takes this dependence into account should be used.
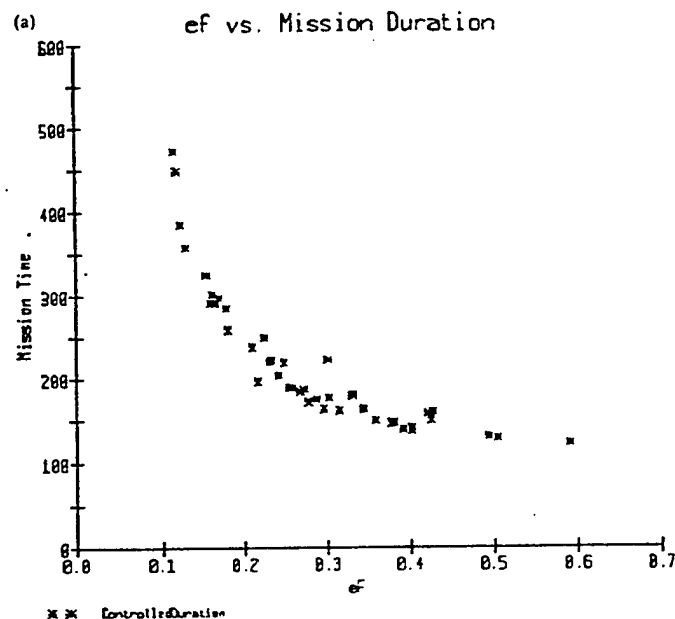
Figure 5   Battle goals plotted against ef

6.   DETERMINING RELIABILITY GOALS

Our threshold is established. We can now say that a gun/round system is mission effective, meeting all of the battle goal thresholds, if the calculated value of *ef is* greater than or equal to 0.40. It may be argued that we are performing inverse regression, and should use the upper limit of the inverse confidence interval rather than the central value. While this upper limit is readily available, we must remind ourselves that the thresholds of the battle goals are set in a very

conservative framework, and further conservatism may lead to unattainable goals. The decision maker has the upper control limit for ef^* at his disposal.

7. THE SIMULATION MODEL METHODOLOGY IN PRACTICE

We have developed a measure of reliability performance that reflects the impact of ammunition failures on battlefield performance. We propose that the decision maker make procurement, surveillance, and training decisions based on ef^* as follows.

7.1. Procurement

When contracting for the purchase of ammunition component i, with all other components in hand, x_i must be small enough so that the overall ef of the system is above ef^*.

The value of increased reliability in a given component can be approximated from the ratio $(T_i/p_i^2)ef^2$, and procurement choices should be made based on cost and gain in ef.

When procuring a new round type, the threshold reliability performance of the round should be established using the simulation model, and the incremental increase in ef should be measured and compared to the round being replaced.

7.2. Surveillance

Upon testing a stockpiled lot of ammunition, if the defect rates of the tested components lead to an acceptable value of ef, pronounce the lot mission effective.

If rework must be done, the partial derivatives $\delta ef / \delta p_i$; should be used to determine the components which will be replaced or repaired.

If the rework is to be a simple component replacement, the defect rate of the replacement component can be substituted for that of the component to be replaced. Thus, ef can be determined before the rework action is taken and we can determine whether the rework will return the round to mission-effective status.

7.3. Training, Logistics, and Administration

Administrators should set threshold recovery times for the different failures based on evaluation using ef. Decisions about which spare parts to carry on the combatant, which to carry on logistics-support vessels, and which to resupply from shore should be based on this analysis.

Training initiatives should be focused on failures with high values Of $(P_i - 1)p_i$, as these failures will deliver the greatest return in terms of effectiveness per unit time of recovery time reduction.

Administrators should evaluate the economic tradeoff of gun crew training versus improving ammunition components directly using the methods we used to construct Table 1.

8. ANALYTICAL STUDY

This section provides the results of analytical studies completed using the simulation previously described in this report. These results are based on evaluation of data output of the computer simulation of an expeditionary force preparing for an amphibious assault in accordance with doctrine and policy in effect at the time. The Joint Munitions Effectiveness Manuals (JMEMs) were used wherever possible for ammunition effects data. Ammunition function, miss distance, and spotting data collected from the observation post (OP) on Vieques Island, Puerto Rico between January 1990 and February 1996 was used. Data collected aboard ship between January 1990 and August 1993 was also used to develop input parameters.

The battle goal of mission time or mission duration is shown as a function of an input parameter of the simulation. The appropriate values of battle goal requirements and known values of the parameter may be applied to aid in the management or quality decision process.

8.1. Scenarios and Threat

Our sponsor, the Naval Surface Warfare Center, Crane Division, Crane, Indiana, provided us with NGFS scenarios that included fleet composition, target types and locations, and deadlines. The simulation we developed inputs this scenario, then subjects the targets to realistic prosecution. For both the 1991-1993 and the 1994-1996 study the standard marine corps amphibious assault western pacific scenario known as MARCOR 6 was used. These plans were studied to develop the target list and strike forces as well as the time windows of opportunity and time requirements for mission accomplishment.

88

8.2. Data collection

If the process of analytical study is likened to the human body, certainly the simulation is the heart and the data is the blood. The quality of the data I would then liken to the oxygen content of the blood. In the case of Navy gun ammunition and gun systems we are low on blood and the oxygen content will barely keep the patient alive.

Reliability data on the components of the MK45 gun either are collected directly by observation aboard the ship as it is shooting qualifications or it is reported by the ship in the routine of doing maintenance or repair of the system. In the case of ship reporting, on the shore side this requires updating of data bases which is reliant on funding levels. In the case of the MK45 and 5 inch ammunition many times the data for these systems has been left not updated for years. That is why this study relies on data by ships observers. Coordination and funding for such an endeavor is difficult therefore data is scarce and decision makers must be informed and take into account the limits of the analysis.

During the period January 1990 to August 1993 gun system reliability information, powder reliability information, gun cycle time information, and repair time information was collected aboard ships performing qualification exercises at the Atlantic Fleet Weapons Training Facility, Vieques Island. This information represents a majority of the input for the NPS simulation. This data was used for both analysis since no additional on board data was available for the latter period.

8.3. Gun System Reliability

The MK45 gun system reliability values that are used in this report are generated from the 1289 rounds shot during successful NGFS qualification that were monitored by a shipboard observer during the period January 1990 to August 1993. They represent those MK45 ships that successfully completed their qualification during the time allocated. This sample most closely represents the state of readiness of ships deployed and in a mission ready state. At each casualty, the ships crew was monitored and the time to repair was recorded. The mean time to repair for each of the reliability blocks is the result of the collection and aggregation of this data.

8.4. Powder Charge Reliability

During 1991 to 1993 , 6871 rounds of ammunition were fired from both MK42 and MK45 5 inch gun systems. Both HE rounds and Puff rounds were shot. Out of these 6871 rounds there were 7 powder casualties. This represents a failure rate of .102 percent. During 1994 to 1996 3788 rounds were fired during qualification during which there were 5 powder casualties. The latter failure rate

89

is .132 percent. Data on powder charge reliability is very good. A failed powder almost always results in communication of that fact to the OP.
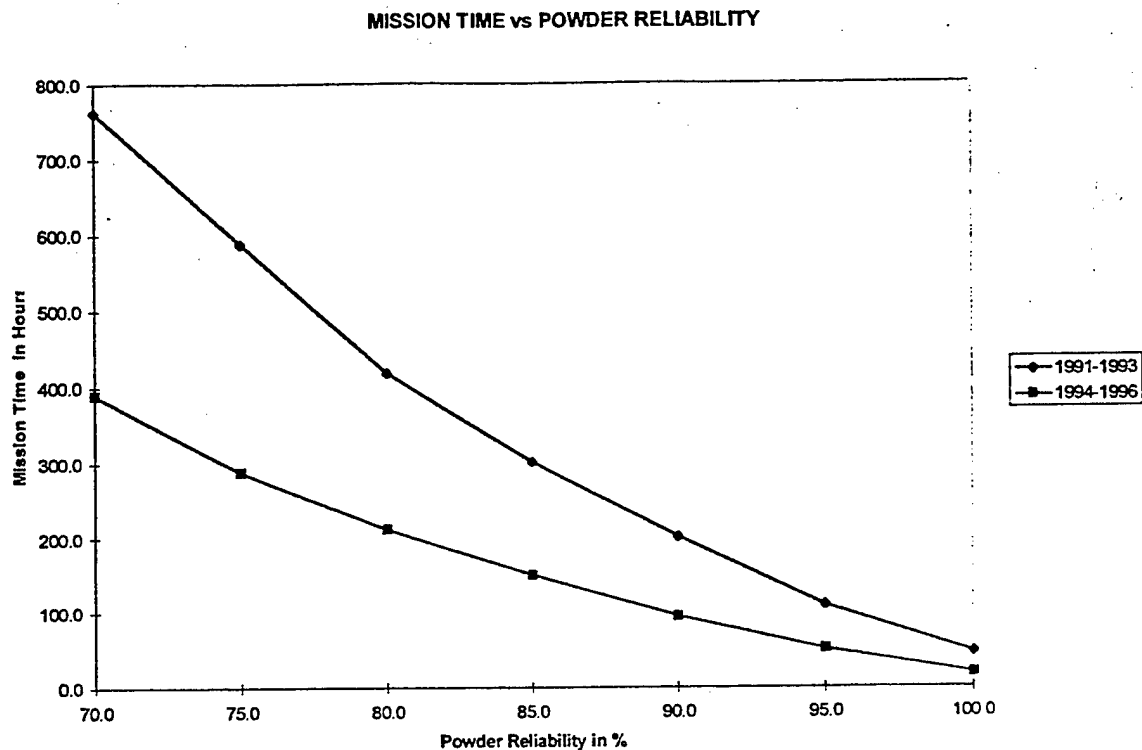
MISSION TIME vs POWDER RELIABILITY

Figure 6   Mission Time vs Powder Reliability

8.5.   Projectile Reliability

There were 6871 observed HE projectiles shot during 1991 to 1993. During that period of time 149 duds and lost rounds were observed. The calculated failure rate of 2.169 percent was used as input to the simulation for projectile failure rate. For 1994 to 1996 there were 20 of 3877 rounds for a rate of .528. These failures are not as obvious as the powder failure but with astute observers the accuracy of this data is very high. With the advent of the high use of GPS in navigation and the automation of IV calculation into the fire control system, lost rounds are almost always projectile failures rather than training errors.
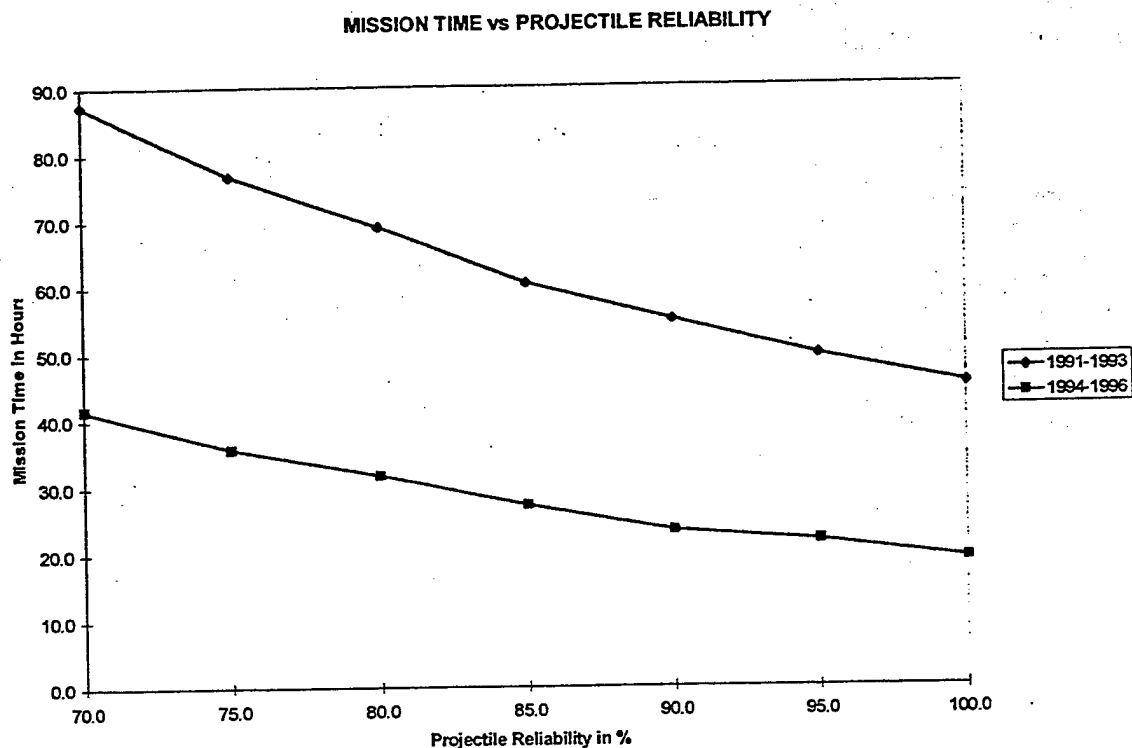
Figure 7   Mission Time vs Projectile Reliability

8.6.   Cycle Time

The cycle time of the gun system is how fast the gun and crew system can deliver ordnance on target? It was obvious from the beginning that the gun was not used to its Design rate of 20 rounds per minute during NGFS qualification exercises. It takes time to get the navigation plot set and spot the gun onto the target. This additional time would make ef calculations using the 20 RPM rate as the base measurement inaccurate.

Cycle time is calculated using only those exercises where the action of the gun is not timed or constrained in any by the nature of the exercise. The d-day exercise, time for spotting rounds , or exercises where the ship is given an arbitrary check fire are not used for the calculations.

Cycle Time is the cumulative time intervals between shots from the first shot of an exercise until the last shot of an exercise, divided by the number of intervals. Cycle Time is represented by the following equation:

$$\text{Cycle Time} = \frac{\text{Total on-line time intervals}}{\text{\# of time intervals}}$$

For the 1991 to 1993 analysis the calculation was 116.6 seconds and for the latter analysis it was 89.5 seconds.

8.7. Spotting Time

The model was built to use a spotting time when it plays against the scenario of targets. The spotting time is defined as the time it takes for the spotters to report back to the ship the fall of shot in relation to the intended target. This spotting time is set in the simulation at the average value of those spotting times observed and recorded aboard ship. For the sake of these analysis it was a constant.

8.8. Miss Distance

The numerical values for miss distance used to perform the analysis to generate this report was the result of a previously reported study titled Validation of Fall of Shot Distribution" dated 15 November 1992. The standard deviation used for the azimuth dispersion was 44.92 meters and the standard deviation for the range dispersion was 68.14 meters for the 1991 to 1993 These values are based on the observation of 423 rounds shot at target number 7 at the Vieques Gun Range. For the 1994 to 1996 study a sample of 17 rounds from 3 platforms was available for dispersion of 12.91 meters and 20.70 meters respectively.

In late 1993 the exercises for NGFS qualification were changed. When this occurred the use of target #7 decreased to almost nothing. The impact of that switch is that there is very little data to validate fall of shot parameters. The use of GPS has certainly reduced navigation error and the use of velocimeters has also reduced the error budget. It is this improvement in single shot accuracy that has enabled the number of spotting rounds to decrease and the cycle time to decrease. Limited data however should be taken into account in using this analysis.

8.9. Graphical Results

To Be Constructed

8.10. Trends

accuracy much better with GPS
accuracy much better with velocimeter
reliability better with loss of old ships
training better with hi tech ways of doing business
data collection slipping fast

STATE OF THE ART

	1991-1993	1994-1996
Sigma x	44.92	12.91
Sigma y	68.14	20.70
number of Registration rounds	5	3
Powder Reliability	99.898	99.868
Projectile Reliability	97.831	99.472
Mission Time in Hours	47.6	19.6
Mission firing rate	.23	.29
Total mission rounds	3900	2040
Mission rounds per gun	650	340
Mission rounds per target	390	204

8.11. Magazine capacity constraints

The constraints on the amount of ammunition initially in the gun magazine or the level of ship fill comes into play for the scenario addressed. If you make the assumption however that the battle group will have ammunition replenishment ships available for underway replenishment of ammunition then the analysis takes this into account quit easily. With the scenario addressed, there would occur at most one replenishment per ship. The state of the art in 1994-1996 is that 340 rounds per ship would be required. If the capacity did not exceed that number, then one hour per gun plus transit time would need to be added to the

estimated scenario time. In the 1994-1996 study that would increase the 19.6 hour scenario time to about 28.6 hours.

9.   SUMMARY

GPS greatly improved navigation picture

Velocimeters greatly improved systemic gun errors

Training seems to have improved cycle times

Collection of data has slowed to a trickle.

Collection of data critical to the quality of the studies.

10.   NEEDS

ship riders and funding to get gun system reliability dat

complete accurate data sheets from observers and graders at OP

Updated scenarios to reflect current threat and current requirements

Next level analytical tools to evaluate integration of other forces and systems in performing a coordinated mission. New capabilities of next generation ammo.

11.   MAJOR REGIONAL CONFLICT WEST

The bosses new mission statement. Study must be upgraded to take it into account for requirements.
Must be done to answer the bosses question, justify budget needs, system needs

12.   HIGH LEVEL ARCHITECTURE

Mandated standards that all analysis models must meet now or very soon. Model must be updated or replaced to meet this.

13.   RECOMMENDATIONS

Support NSS effort to upgrade the model and incorporate MRC West. If not NSS then whatever HLA compliant model effort that takes its place and is blessed by the boss

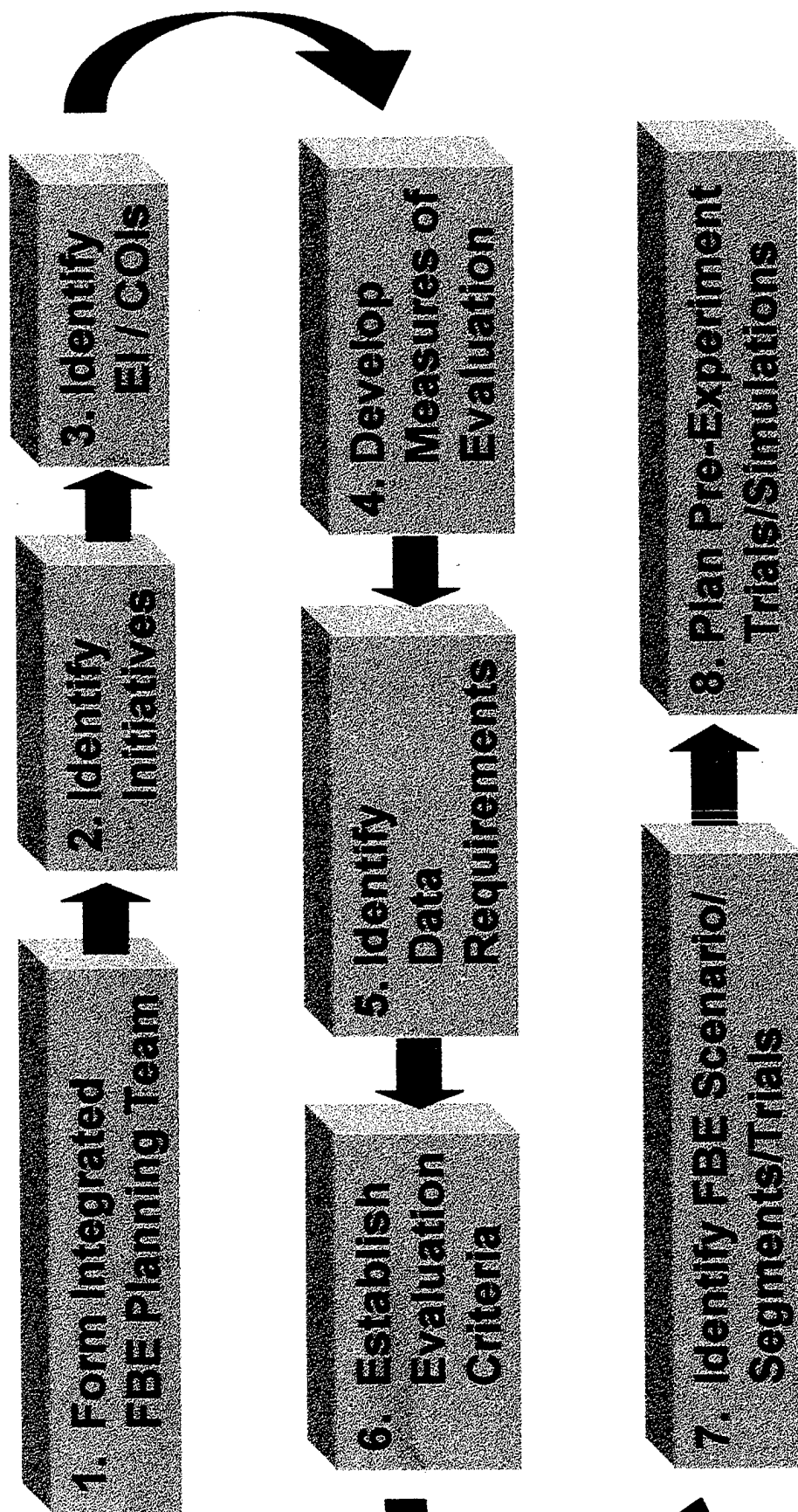Get data collection back on track

14. REFERENCES

[1] Bailey, M., Bartroli, M., Callahan, A., and Kang, K., *Establishing Reliability Goals for Naval Major-Caliber Ammunition*, Naval Research Logistics, Vol. 39, pp. 877-892, 1992.

[2] Bratley, P., Fox, B., and Schrage, L., A *Guide to Simulation* (2nd ea.) SpringerVerlag, New York, 1983.

[3] *Training and Qualification of Ships in Naval Gunfire and Shore Bombardment (Gunsmoke Manual)*, COMNAVSURFLANTINST 3570.2D, August 1989.

[4] *Joint Munitions Effectiveness Manual (Surface-to-Surface), Effectiveness Manual for 5-inchIS4 Naval Single Mount Mk45 with Gun Fire Control System Mk86*, (Unclassified), Vol. 2, October 1981.

[5] Kleijnen, J.P.C., *Statistical Techniques in Simulation, Part 1*, Marcel Dekker, Inc., New York, 1975.

[6] *Material Readiness Data Base User's Manual, Version 111.1*, U.S. Navy, NWAC, Corona, CA.

[7] *MODSIM 11 Reference Manual*, CACI Products Company, LaJolla, CA, 1990.

[8] Sargent, R.G., and Shanthikumar, J.G., *"A Unifying View of Hybrid Simulation/ Analytic Models and Modeling,"* Operations Research, *31(6)*, 1030-1052 (1983).

[9] *Weapons Systems Fundamentals, NAVORD OP 3000, Vol. 2, Analysis of Weapons*, Published at the direction of Commander, Naval Ordnance Systems Command, 1971.

Institute for Joint Warfare Analysis

Process for Determining Measures of Evaluation

MOE Process-THH
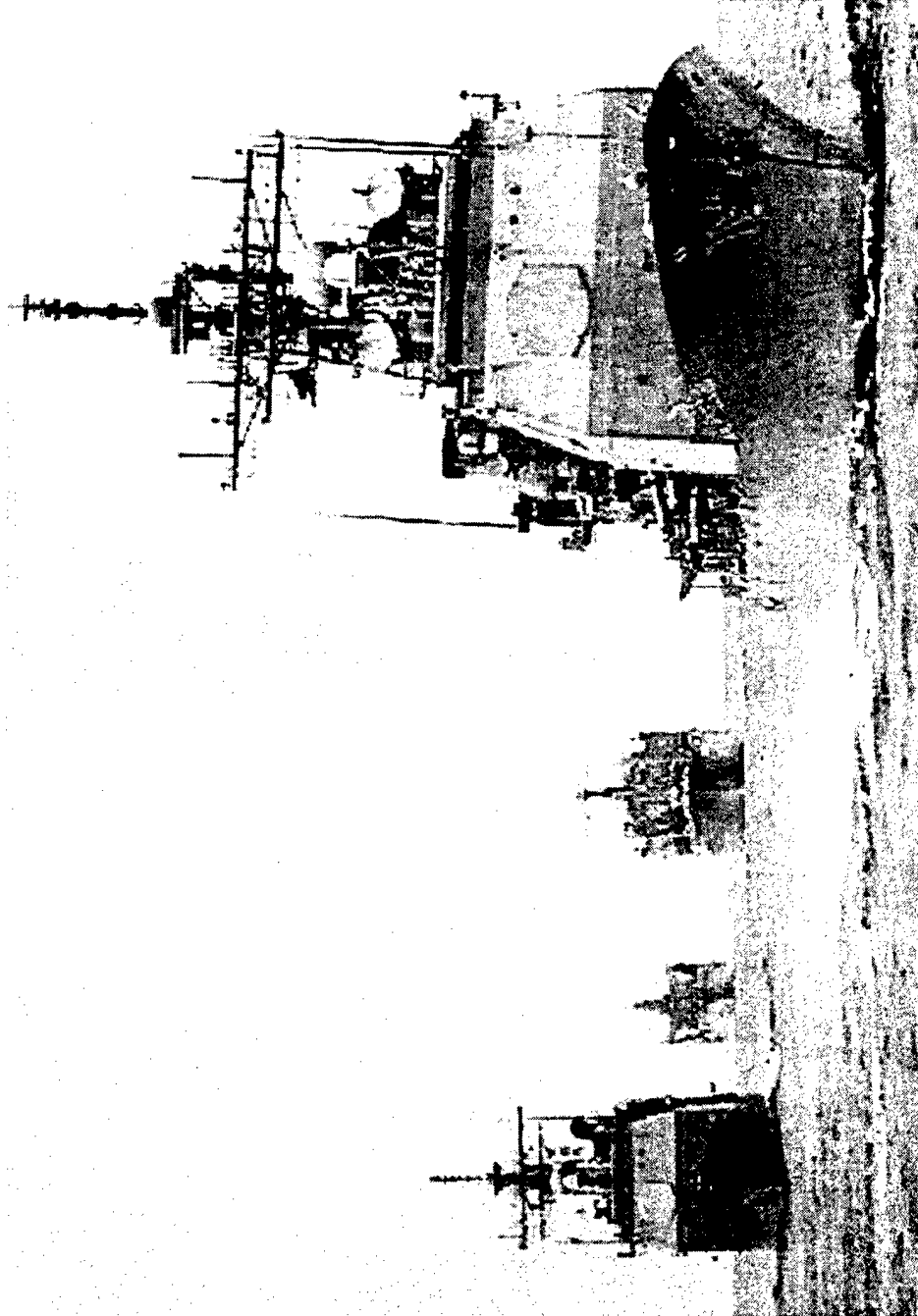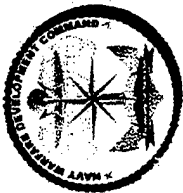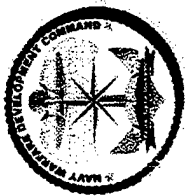
Initial FBE Planning Steps

1. Form Integrated FBE Planning Team

2. Identify Initiatives

3. Identify EI / COIs

4. Develop Measures of Evaluation

5. Identify Data Requirements

6. Establish Evaluation Criteria

7. Identify FBE Scenario/Segments/Trials

8. Plan Pre-Experiment Trials/Simulations

1. Form Integrated FBE Planning Team

Should Include Representatives from:

- Fleet Commanders
- Data Collection and Analysis Team
- CHENG
- System/Process Developers
- Other Initiative Developers
- Red Team
- War Gamers / Simulators
- Others?

MOE Process-THH

2. Identify FBE Initiatives

- Include all Proposed Initiatives that are possible candidates for FBE Evaluation, such as:
  - **Systems, Process, Architectures, Tactics and Doctrine**

- Select those Initiatives that can best be evaluated in an FBE structure and environment.
  - Other Options for Evaluation include:
    - **Simulation, Modeling, DT&E, OT&E, Exercises**

- Include Fleet Initiatives if possible
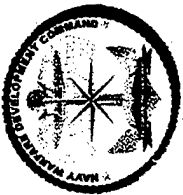
3. Identify *Experimental Issues (EI)* or *Critical Operational Issues (COI)*

- Some Issues may be provided by CINCs, Fleet Cdrs, PM/PEOs

- Other Issues/Sub-Issues may be formed by <u>identifying</u>:
    - The Functions & Capabilities which must be performed by the System, Process or Architecture (Dendritics Help)

- Select those that may have the most effect on mission accomplishment

- Issues are stated as <u>Questions</u> to be addressed or resolved.
    - "Is the JFMCC planning process satisfactory for Joint Operational Mission Planning?"

5

100

Functions and Capabilities Dendritic
(Tool for Deriving Issues & Measures)

Functions→Issues
Capabilities→MOEs

Level Two
Capabilities

Level One – Functions

CMD/Control
- Chat
- IVOX
- Links
  - Link 11
  - Link 16
- :

ISR
- UAV
  - UAV 1
  - UAV 2
- JSTARS
- Lantirn
- Hairy Buffalo

JFMCC Planning
- AAW
- ASW
- MIW
- :

Time Critical Targeting
- LAWS
- Mensuration
- BDA
- :

:
:

Joint Operations Functions & Capabilities
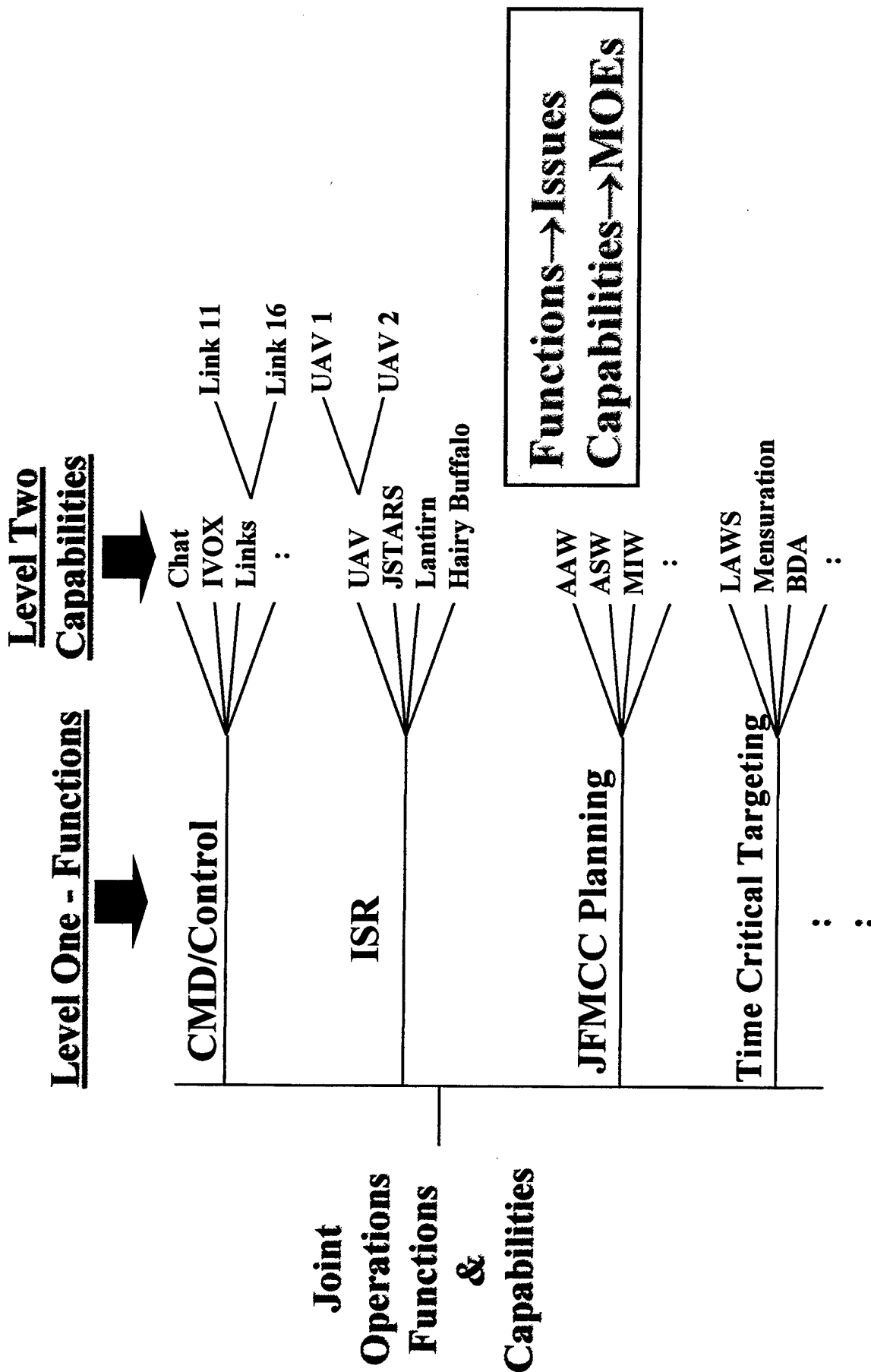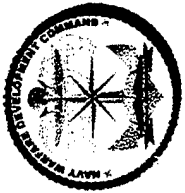
4. Identify Measures of Evaluation

- Identify the measures to be used for analysis, evaluation, & determining the worth of the function

- Measures of Effectiveness (MOE) & Measures of Performance (MOP) Quantify the Results of the Experiment

- COIs/EI's can be Answered by a Single or Multiple MOEs & MOPs or Combination

- The Measures should drive the FBE Design, Data Set, Resource Requirements, Analysis & Reporting

- Measures are usually system or concept unique

7

Measures of Effectiveness (MOE)

- *A MOE is a measure which expresses the extent to which a combat system accomplishes or supports a mission or task. Usually a System Function Capability. Designed to address an Issue. (May vary with Level & Scope of System)*

Examples:

- ## Radar System MOEs:
  - Detection, Identification, Single Target Track, Situational Awareness, Track While Scan (Capabilities)

- ## FBE EI & MOEs :
  - Augmented C2: IVOX, Chat Capability
  - TCT EI: Mensuration Capability
  - JFMCC EI: Planning Capability
  - COPS: Situational Awareness Capability

103

8

Measures of Performance (MOP)

- *A MOP is a quantitative or qualitative measure of a system's capabilities or specific performance function. (Rates, Ratios, Percents)*

- <u>RADAR MOPs:</u>
  - Average Detection Range of a 2 Square Meter Target
  - False Alarms Rate
  - Ease of Operation of Controls and Displays
  - Mean and Variance of Detection Ranges

- <u>COPS/MOP:</u> Percent of operators judging situational awareness to be unacceptable

- <u>Mensuration MOP:</u> Average Time to Mensurate Targets

Data Requirements are Defined to Answer Each MOP

Specific FBE Events are Devised to Provide These Data
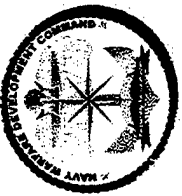
5. Identify Data Requirements

- **Data Requirements (DR) are derived directly from the MOPs**

- **Documentation of Data Sources, Type and Format should be included in the FBE plan**

- **Data Type, Fidelity and Amount of Data must be Coordinated with Operator & Instrumentation Personnel**

- **A Data Requirement Dendritic May Help in Identification of Appropriate Data**

- **Additional Data that could influence the results are identify bias should also be collected ("background data")**

Traceability Dendritic for COI/MOE/MOP/DR

(From TEPRS)

<u>Targeting Capability?</u> (EI/COI)

- ▲ Target Search (MOE)
  - ⊟ ▲ Search Rate (MOP)
    - ▲ Indication that Assigned Sector was Scanned (Y, N)
    - ▲ Start Operations Time
    - ▲ Stop Operations Time
- ▲ Target Production ☐ ☐
- ▲ Target Tracking
  - ⊟ ▲ **Location Accuracy**
    - ▲ Average Range Error
      - ▲ Actual Location (X, Y, (Z) coordinates) of Firing Unit
      - ▲ Actual Location (X, Y, (Z) coordinates) of the Target
      - ▲ Reported Location (X, Y, (Z) coordinates) of Target Unit
      - ▲ Reported Location (X, Y, (Z) coordinates) of Fire Unit

Blue   – MOE
Green – MOP
Red   – Data Rqmt

11

106

Baseline Correlation Matrix
Correlates EI/Criteria/MOE/MOP/DR

Baseline Correlation Matrix

EI (COI)	Criteria	MOEs	MOPs	DR
Experiment Issue 1: Is the Time Critical Targeting Process satisfactory for the amphibious assault mission?	Criterion 1.1 The TCT process from time of Target Detection to time of Weapon Firing should be less than 10 minutes.	MOE 1.1.1 Target Search Capability	MOP 1.1.1.1 Search Rate	Start Operations Time Stop Operations Time Indications that Sector is Covered Area Covered
		MOE 1.1.2 UAV Target Detection Capability	MOP 1.1.2.1 Average Range of Detection MOP 1.1.2.2 Proportion of Detections	Target Location UAV Location at Detection Number Available Targets Number of Targets Detected
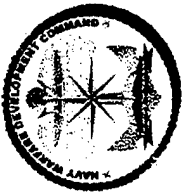
6. Establish Evaluation Criteria

- **Evaluation Criteria Are Standards Used to Evaluate the Mission or Operational Impact of an Experimental Issue or System.**

  - **Stated Qualitatively or Quantitatively**

  - **Must be Unambiguous, Testable & Measurable**

  - **Helps with insights for judging merit of concepts**

  - **Should include conditions under which it applies**

- **Tied directly to Experimental Issues, Operational Requirements, or System Program Needs**

- **Each Criterion should be correlated with at least one MOP**

13

108

Evaluation Criteria Alternatives

1. **New System/Architecture versus Replaced System or Architecture**

   – **Use Old System as Baseline Using <u>Same Experiment</u>**

2. **With New System/Architecture versus <u>Without</u> System/Architecture**

   – **Measure Performance of an Organization With the New System and Without the Systems (EW - RIL)**

3. **New System versus Predetermined Standard**

   – **Compare System Performance with a Standard or Requirement**

4. **Evaluate New System/Architecture using SME's**

   – **Least Desirable, due to possible SME bias, but may be only method available**

14

109

MOE Process-THH

Examples of Evaluation Criteria

Specific Requirements (Easier for comparison) i.e.

- Time Critical Targeting cycle will be < 10 minutes from initial detection until firing.

- BDA will received by the Battle Watch within 15 minutes

- The Mensuration cycle will be completed within 4 minutes

- The mine neutralization vehicle will have a 90% success rate.

- The JFMCC Planning Cycle will be completed within 5 hours.

- The COPS will "significantly" enhance Battle Watch Situational Awareness   (Problem: How to Evaluate?? What are MOPs ???)

15

110

Institute for Joint Warfare Analysis

7. Identify *FBE Scenario/Segments/Trials*

- **Although overarching scenarios will vary from Fleet to Fleet, it is the segmenting & structuring of various trials (within the scenario) to identify factor effects which is most important!**

- **Experimental Segments should be structured to address the COIs/EIs, Objectives, and MOPs.**

- **Specific Controllable Trials or Subset Experiments should be identified that can evaluate major factor contribution to Mission Accomplishment**

- **Tradeoff is between "Controlled Scenarios" (lose realism) and "Total Freeplay" (factor confounding limits evaluation capability)**

111

Developing Experiment Conditions & Scenario Segments

- **Identify Critical Operational Factors for Evaluation**
- **Establish conditions needed for determining factor effects**
  - **A Factor and Condition Matrix may help**
- **Breakdown Scenarios into specific segments or experimental trials**
- **Red Team varies conditions in semi-controlled manner w/o seriously violating realism**
- **Systematic variation allows important factor effect insight and knowledge**
- **Segmented Trials can be interwoven into FBE w/o being known to fleet operators.**
- **Red Team Play is Vital to obtaining Credible Data**

Example: Factors & Conditions for FBE

FACTORS	CONDITIONS
JFMCC	Without CWC Input With CWC Input
Command & Control	Centralized, Decentralized
Comm. Jamming	On and Off
ISR	JSTARS, Lantirn, Hairy Buffalo
Comm. Augmentation	Chat, IVOX
Visibility	Day, Night, Fog
UAV	Two Types
Fire	Air Wing, Cruise Missile
MCM	RMS, Other
Targets	SCUD, Artillery, Troop Movement
Enemy Mission	Defend Attack

Example: EI/COI Factor Design Matrix

EI Factors (Variable)	Control	Factor Conditions (Levels)
Command and Control	Systematically varied	Centralized, decentralized
Comm Jamming	Systematically varied	Jamming On – Off
Augmented Communications	Systematically varied	Chat, IVOX
Fire support mission	Systematically varied	Air Wing, Cruise Missile
Firing	Systematically varied	Live, Simulated (Model V&V)
Missiles	Systematically varied	Current, Future
Combat intensity	Systematically varied	Supported, surge, and peak
Ground Target Threat	Systematically varied	SCUD, Artillery, Troop
JFMCC Planning Methodology	Systematically varied	Method 1, Method 2 (W-W/O CWC Input)
BG Movement	Tactically varied	IAW with scenario
ISR	Systematically varied	Remote, Lantirn, P-3
UAV	Tactically varied	Live, simulated
UAV Tactics	Systematically varied	IAW Tactics
MCM	Systematically varied	RMS, Current, Future
Mine Threat	Held Constant	Minefield not varied 3 types
Software	Systematically varied	Version 2.1, 3.0
Tactical Organization	Held Constant	Fleet Specified
Doctrine	Systematically varied	IAW specified Scenario
Personnel - Ship action	Uncontrolled	As occurs
Weather	Uncontrolled (Measured)	As occurs Day and night
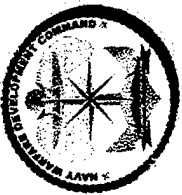
8. Plan for Pre-Experiment Trials

Why Have Pre-Experiment Trials?

- Allows training of operators and data collectors

- Ensures experiment procedures give desired system results

- Ensures experiment procedures are efficient and accurate

- Confirms that necessary data is being collected

- Ensures that data collection does not interfere with system operators & experiment realism

- Allows computer and system shakedown

- Provides an opportunity to generate a baseline for comparative analysis

Note: Can Use Simulation with a Mini-Scenario or War Game for Pre-Experiment Trials

115

20

Institute for Joint Warfare Analysis

Experiment Observations

- **Experimental Issues should be well defined and limited and established early.**

- **COI-EI/MOE/MOP/DR should be traceable**

- **Instrumentation and Data Collection should be available at critical nodes.**

- **Scenarios Segments and Trials should be structured for analysis & evaluation when possible after <u>Initial</u> Fleet Scenario is established**

- **Concepts should have baselines for comparison if possible**

- **Criteria should be unambiguous for evaluation**

MOE Process-THH

Operational Fires MOPs/Data Elements

OP 3 Employ Operational Firepower

M1	Minutes	To attack target after most recent information on target provided.
M2	Percent	Of HPTs successfully attacked.
M3	Percent	Of missiles, rockets, and other long range attack systems successfully engaged targets.

OP 3.1 Conduct Joint Force Targeting

M1	Hours	To develop attack plan after identification of HPT.
M2	Hours	To issue CJTF's HPT categories (after CINC's Warning Order).
M3	Hours	To issue CJTF's Prohibited Target Guidance (after CINC's Warning Order).
M4	Hours	To issue FSC Measures Guidance (after CINC's Warning Order).
M5	Hours	To pass JTCB Guidance to targeting agencies (e.g., JFACC) (before ATO-cycle begins).
M6	Hours	To produce Joint Force Commander apportionment guidance (after CINC's Warning Order).
M7	Percent	Of desired results achieved (by expected conclusion of given phase or time line).
M8	Percent	Of desired results achieved by theater operational firepower (within specified time/phase).
M9	Percent	Of JTCB target priorities differ from CJTF, CINC and NCA Guidance.
M10	Percent	Of selected targets for which accurate coordinates available.
M11	Percent	Of JTCB selected targets reviewed for political ramifications.
M12	Percent	Of targets susceptible to non-lethal kill allocated to non-lethal attack systems.
M13	Percent	Of enemy NBC delivery systems targeted by friendly forces.

OP 3.2 Attack Operational Targets

M1	Minutes	To get ordnance on target after initiation of task
M2	Percent	Execution of missions requested by components.
M3	Percent	Of high priority missions executed within specified time.
M4	Percent	Of maneuver forces secure assigned objectives.
M5	Percent	Of missions flown/fired achieve desired target damage.
M6	Percent	Of operational fires on time in support of maneuver forces.
M7	Percent	Of preplanned targets successfully attacked during operation.
M8	Percent	On time of missions with given times on target.
M9	Percent	Of enemy NBC delivery systems engaged/destroyed by friendly forces.

117

MOE Process-THH
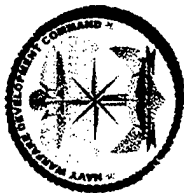
Tactical Fires MOPs/Data Elements

NTA 3 Employ Firepower

M1	Percent	Of high priority targets (HPTs) successfully attacked.
M2	Percent	Of missiles, rockets, etc., successfully engage targets.
M3	Percent	Of higher authority tasked missions accomplished.
M4	Percent	Of desired results from attacks or engagements.
M5	Percent	Of missions assigned by higher authority are successful.
M6	Percent	Actual weapons used compared to projected.

NTA 3.1 Process Targets

M1	Percent	Of desired results achieved by expected conclusion of a given phase or time line.
M2	Percent	Of selected targets have accurate coordinates available.
M3	Percent	Of targets susceptible to non-lethal kill allocated to non-lethal attack systems.
M4	Time	To identify target as HPT.
M5	Hours	After receipt of Orders to identify High Priority Targets.
M6	Hours	After receipt of Orders to review Prohibited Target Guidance.
M7	Hours	After receipt of Orders to review FSC Measures Guidance.
M8	Hours	Before ATO-cycle begins, JTCB Guidance is passed to targeting agencies (e.g., JFACC).
M9	Hours	For the targeting cycle to be completed.
M10	Number/day	Targets administratively processed during a given phase or time requirement.

NTA 3.1.1 Request Attack

M1	Time	In advance of attack targeting strategy is established.
M2	Percent	Of attack requests submitted in compliance with a given phase or time requirement.
M3	Percent	Of targets correctly identified and located.
M4	Time	To request attack after targeting priority established
M5	Percent	Of requests forwarded with all required data (i.e. Target ID and location).
M6	Time	To develop and issue request.

23

118

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center 2
   8725 John J. Kingman Road, Suite 0944
   Ft. Belvoir, VA 22060-6218

2. Dudley Knox Library 2
   Naval Postgraduate School
   411 Dyer Road
   Monterey, CA 93943-51013

3. Research Office Code 09 1
   Naval Postgraduate School
   Monterey, CA 93943-5138

4. Superintendent Code 00 1
   Naval Postgraduate School
   Monterey, CA 93943

5. Institute for Joint Warfare Analysis Code JW 10
   Naval Postgraduate School
   Monterey, CA 93943

6. VADM Dennis McGinn 1
   Deputy CNO for Warfare Requirements and Programs (N7)
   The Pentagon
   Washington, DC 20301

7. VADM Arthur K. Cebrowski 1
   President
   Naval War College
   686 Cushing Road
   Newport, RI 02841

8. RADM Robert G. Sprigg 1
   Commander
   Navy Warfare Development Command
   686 Cushing Road
   Newport, RI 02841

Naval Postgraduate School
Monterey, California