Department of the Navy
Small Business Innovation Research (SBIR)

Guidebook to SBIR Experimentation

Guidance for Small Business Innovators, Naval Subject Matter Experts,
and Program Managers on Using the SBIR Vehicle
to Realize Experimentation Objectives

Version 1

March 2021

DISTRIBUTION STATEMENT A
Approved for public release: distribution unlimited.

Record of Changes

Revision Number	Date Published	Summary of Changes
1	3/17/2021	New Version

Contents

List of Figures

List of Tables

List of Appendices

1 Executive Summary

Organizations within the U.S. Navy have benefited from the use of experimentation in closing warfighter gaps, reducing risk, and validating innovative technologies. Many of these organizations and their Technical Points of Contact (TPOC), Project Managers (PM) and subject matter experts (SME) have utilized practices and tools that have helped them navigate the acquisition process and succeed in their efforts. Additionally, small businesses and Naval Program Managers can use experimentation to gain exposure for their innovation, receive fleet feedback and in-phase developmental feedback prior to acquisition testing and evaluation.

This guide aims to demystify experimentation processes and provides a basic introduction into experimentation including considerations for types of experiments and venue selection, functional areas to address and who to work with in each, and finally an overview of the process for planning, preparing, executing, and closing out an experiment. This guidebook seeks to help decipher the engineering and operational challenges that small businesses face in experimentation and help align their environment or event for initiative. It also includes the best practices from naval doctrine and for evaluation of required resources. The purpose is to serve as a comprehensive reference tool for SBIR community innovators in Phase II or III of maturity. Consult Figure 1-1 for an overview of the award structure and milestones associated with each phase.

SBIR projects can provide quality first candidates as many of those projects are ready for the next step of exposure to the fleet. Science and Technology (S&T) experimentation is an effective method of increasing this exposure while also ensuring operational relevance. S&T can enable rapid transition of capability to the warfighter and is a shared objective for the PEO, NAVWAR, and Department of the Navy (DoN) as a whole. Experimentation and analyses provide a feedback loop to the PEO in support of strategic vision and direction, helping to close knowledge and requirement barriers.

While the experimentation route can appear long and fraught with complexity, there are agencies and mentors exist to facilitate and mentor the community through these processes. One of the goals of this effort is to highlight the timeline for participation in experimentation so that both the SBIR small business and their government PM/TPOC/SME can make sure that the required resources (time and money) are built-in to the SBIR Phase II/III project as soon as possible. The recently established DoN SBIR Experimentation Cell (SEC) is also available to facilitate, mentor and train both the small business and government personnel for participation in experimentation events.

As an introductory reference guide for Department of Defense (DoD) users, current DoD policies are complimented by capturing and consolidating approaches, best practices, and recommendations. Additionally, it provides the reader with discretionary expertise that can be tailored to the individual experiment. As a dynamic document, this guide will undergo cyclical reviews and revisions to ensure the latest updates are implemented. This work is not intended to be used in lieu of official policy or training, but instead like a simplified study guide, as it references a multitude of process guides, organizational procedures, doctrinal guidance, and informed best practices to create a tailored generalized SBIR Communicant reference manual.

Approved, DCN# 43-8003-21

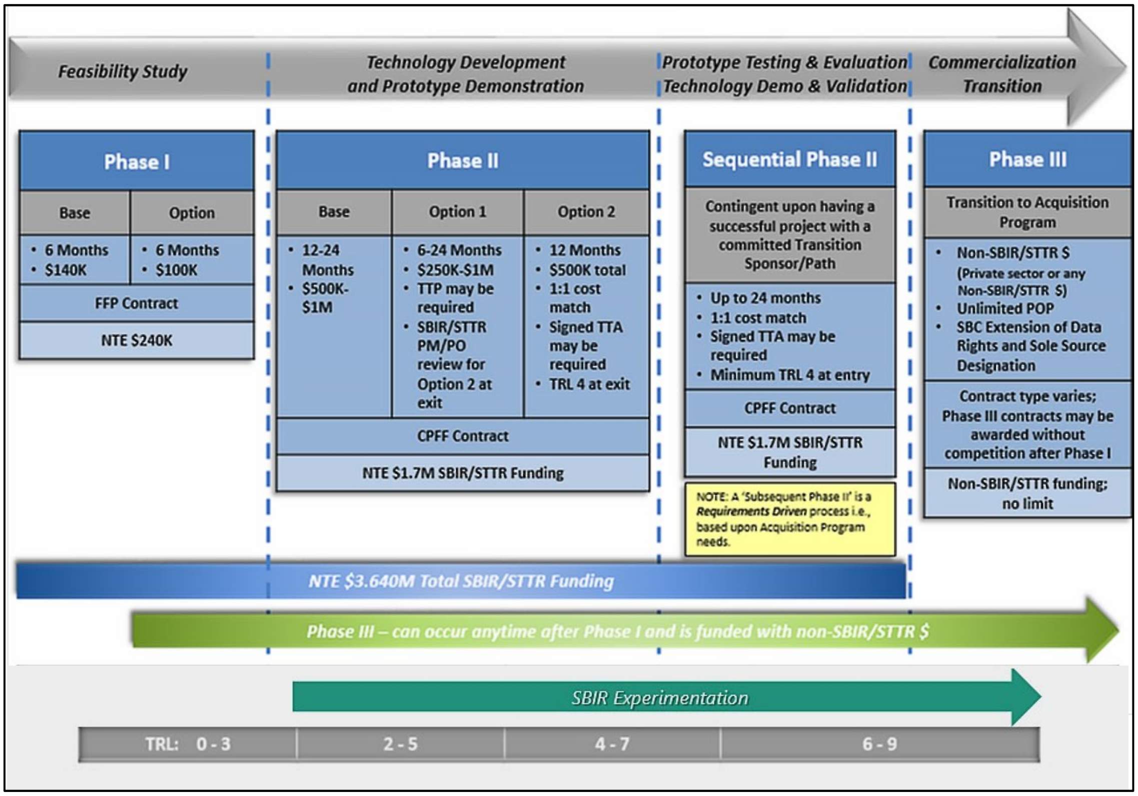

FIGURE 1-1 SBIR PROGRAM AWARD STRUCTURE

2 Introduction

Naval leadership has recognized the need for innovation in an era where national warfighting technical dominance is eroding as adversaries exponentially increase their capabilities.[1] The nation's near-peer competitors have been studying military capabilities and are gaining a competitive advantage and exploiting vulnerabilities.[2] In order to stay relevant in the high-technical battle for supremacy, naval forces need to provide relevant and scalable answers to warfighter gaps.

Experimentation is a key enabler used by naval acquisition and engineering commands to shepherd warfighters into the future by discovering and examining innovative and technically advanced solutions to solve warfighter gaps. One acquisition strategy is to target the Navy SBIR programs. These programs are used to deliver small business innovative technology solutions to today's warfighters. SBIR's mission is accomplished by applying the agility, dedication, and ingenuity of small business entrepreneurs to the Research and Development (R&D) needs of the Navy.

The SBIR community supporting this innovation effort is comprised of small businesses, Systems Command TPOC and various supporting acquisition and engineering organizations. Upon deciding to invest in SBIR technologies, technical capabilities provided by small businesses must be efficiently and securely installed and implemented. The SBIR program has the statutory purpose to strengthen the role of innovative small business concerns (SBCs) in federally funded research or research and development (R/R&D). Specific program purposes are to: (1) Stimulate technological innovation; (2) Use small business to meet Federal R/R&D needs; (3) Foster and encourage participation by socially and economically disadvantaged small businesses in working in technological innovation; and (4) Increase private sector commercialization of innovations derived from R/R&D, thereby increasing competition, productivity, and economic growth.

[1] The Defense Innovation Initiative, Secretary of Defense, 2014
[2] "A Design for Maintaining Maritime Superiority", Chief of Naval Operations, 2018

Traditional SBIR sponsors and their industry partners, while experts in their own fields, often lack in-depth process knowledge on how to conduct technical experiments on ships, aircraft, and naval shore bases. A lack of breadth and depth of knowledge can adversely impact the ability to plan and execute experiments on naval assets efficiently and effectively. Furthermore, stakeholder complexities, extended timelines, and increasing resource costs to meet requirements for installation of Command, Control, Communications, Computers, Combat Systems, Intelligence, Surveillance and Reconnaissance (C5ISR)-related hardware and/or software on U.S. Navy (USN) and Military Sealift Command (MSC) platforms and shore commands hinder Navy innovation and S&T rapid development. Cybersecurity and certification for authority to operate (ATO) are also obstacles to experimentation and implementation that have unique, and often changing, requirements. Failure to understand these requirements and the cost/time to implement them in a timely manner, even if only in an interim state, can prevent participation in an experimentation event and even derail implementation of the system no matter how innovative the technology.

2.1 Doctrinal Alignment

Current doctrine establishes how the warfighter will communicate and pass data (intelligence, surveillance, target acquisition, and reconnaissance) for future engagements. These doctrinal products then produce gaps in the current forces and drives the acquisition process. These gaps are traditionally actionable and assist the warfighter in efficient observation, orientation, and decisions. Warfighting gaps exist when current capabilities or measured performance do not meet stakeholder requirements. A warfighting gap can address a capability requirement as well as the need to advance existing concepts and capabilities under specified conditions. Many military demonstrations utilize Tactics, Techniques and Procedures (or TTP, as it will be referred to throughout) to demonstrate or explain the capability in question under these conditions. It is important to note that in some communities, TTP stands for Technology Transition Plan. This plan is the agreement between the funding sponsor (SBIR) and the acquisition program office to transition a promising technology at completion. For purposes of this guidebook, this type of document will be referred to as a Technology Transition Agreement (TTA) instead. The gap statement (sometimes referred to as the problem statement) must specify the stakeholder organization and describe the shortfall. The statement should be summarized in a single sentence supported by facts and relevant impact.

2.2 SBIR and Experimentation

SBIR, in concert with the SYSCOMs, seeks to provide data to serve as the foundation for addressing warfighter gaps. SBIR and other research sponsors rely on SYSCOM engineering communities and acquisition program managers to obtain necessary approvals for installation of technologies on aircraft, ships, submarines, and more within shore facilities/platforms. Current experimentation programs such as Fleet Experimentation (FLEX) and Trident Warrior (TW) are large-scale annual events that will facilitate installation processes if proposed projects or initiatives are deemed acceptable. There are many types of experimentation events varying in both size and complexity; the maturity of the technology at the time of the event may factor into what type of event is pursued. A prototype will start with small events and testing and gradually grow into larger events as it develops in both maturity and complexity. These events and more are described in detail in future chapters.

SBIR candidates progress through a three-phase guidance process in fostering their innovations. While this guide caters mainly to Phases II and III, all three are described here. Refer to Figure 1-1 for an overview of the award structure.

Phase I, the idea generation phase, establishes proof of concept through a feasibility study. These studies determine the scientific/technical merit of an idea or product that may provide a solution to a USN need. The Base and Option periods of performance are not to exceed six months each. Commercial potential is key to moving from Phase I to Phase II, as further federal support is granted upon determining performance quality. Once granted a Phase I contract, awardees can submit an initial Phase II proposal for consideration and selection.

Homing in on technology and ideas more closely, Phase II is the principal research or R&D effort as well as where much of the funding is spent. It is expected to produce a well-defined deliverable prototype during this phase. The DoN will evaluate proposals using the criteria in Section 8.0 of the DoD Program Solicitation. For an award to be granted, technical merit is the most important evaluation criteria, followed by personnel qualifications and commercialization potential of equal importance. The period of performance is generally 24 months. A Phase II contractor may receive up to one additional, sequential Phase II award for continued work on the project. A 'Subsequent Phase II' is a requirements-driven process within the scope of the original project and is based upon acquisition program needs.

As Phase II is the primary demonstration period, candidates will find that much of this guide dives into testing procedures and may help to decipher the process as prototypes are built. Timelines, opportunities for exhibition, and experimentation basics are just a few of the topics covered to assist in familiarizing innovators as projects progress from the idea to implementation stage. It is important to note that acquisition programs may not receive Milestone B approval until certification per 10 U.S. Code § 2366b.[3] Certification is conducted by the Milestone Decision Authority and takes into consideration the cost, objectives, schedule, and performance of the program.

The final phase in which commercialization and demonstration can finally take place is Phase III. This phase involves maturing the technology for delivery to defense or commercial customers. The goal is to transition a company's SBIR effort into products, tools or services that benefit the Navy acquisition community, and ultimately lead to investment from acquisition program funds and industry.

Many experiments are conducted with technology, in this case SBIR technology, before entering a formal acquisition process. Experiments also assist decision makers in fast-tracking the development of promising warfighter capabilities. An experiment can also identify additional research needed to address the warfighter gap. Finally, experiments enable SBIR technologies to have every possible advantage to impact investment and prototype decisions and further develop already-identified improvements.[4]

3 Experimentation Overview

An experiment is fundamentally an attempt to learn whether a technology (in this case a SBIR technology), a TTP, or a combination of both has the possibility of addressing a warfighting gap. The knowledge gained during an experiment is different from other knowledge in that it is founded on observation and experience.[5]

To conduct the experiment safely and effectively, a myriad of activities must take place. These include designing the experiment (defining the problem statement, selecting the type of experiment, selecting the venue and event, defining support requirements from the operational community, etc.), scheduling resources (operational platforms, manpower, communication circuits, etc.), developing experiment plans and data collection and analysis plans (DCAP), conducting installation and removal of equipment, and numerous others. Considerations for these activities are discussed further in this guidebook.

A note on terminology: The definition and usage of terms used throughout this guide are in the context of experimentation and are based on empirical research and the authors' experience. The term "initiative," for example, has two meanings depending on where and how it is being used. Generally, "initiative" is used to describe a stakeholder's attempt to achieve a goal or solve a problem to fill warfighter gaps. Put simply, the stakeholder has "initiated" a project that will explore a proposed solution to the warfighter gap. Separately, in FLEX or TW events, the initiative is an experiment that has been nominated to or accepted into a formal DoN experimentation process and has been vetted in some way. This guide will use the former definition for

[3] Legal Information Institute
[4] Department of Defense Prototyping Guidebook, 2019, p. 8
[5] Alberts & Hayes, 2002

"initiative": to describe the projects that will use experimentation to complete their objective, prove a thesis, or demonstrate their capability or worthiness to solve a problem.

3.1 Functional Areas

Whether creating an experiment from scratch or participating in a scientific or fleet experimentation venue, the experiment activities fall into four general categories or areas: 1) Operations, 2) Engineering, 3) Analysis, and 4) Administration, Manning, and Logistics. Of special note, these areas also align with the type of personnel expertise needed to address requirements. Each will be generalized in this guide and defined subprocesses will be expanded in the appendices. Noting Figure 3-1 below, these functional areas cross each other and have interdependencies that will be discussed in future sections. Administration, Manning, and Logistics touches all the other areas. Each of the areas also has a cost factor that must be considered.

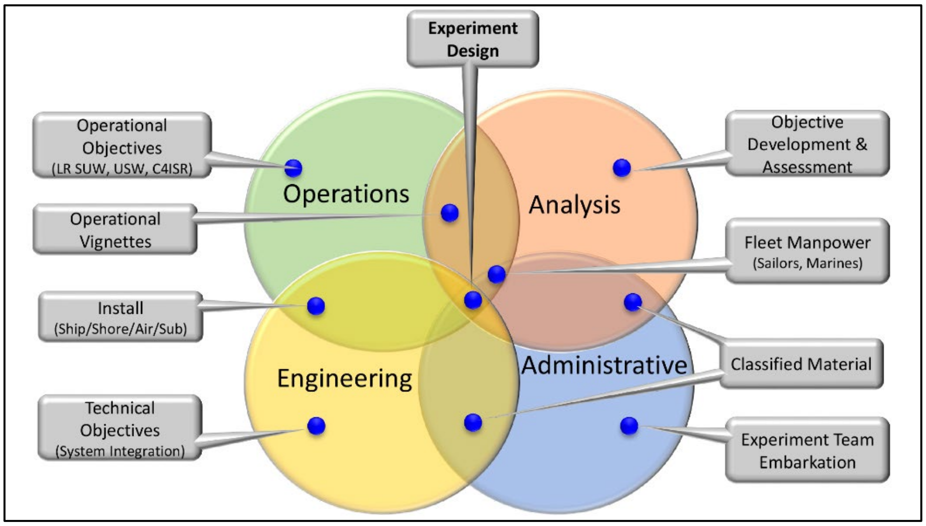

FIGURE 3-1 AREA INTERSECTS & INTERDEPENDENCIES

3.1.1 Operational Areas

Operational areas in an experiment essentially answer the who, what, when, where, and how. *Initiative Planning* is most critical in the successful design and execution of an experiment. This type of planning provides details on proposed experiment objectives, design, timeline, and deliverables to produce tangible products regarding Doctrine, Organization, Training, materiel, Leadership and Education, Personnel, Facilities and Policy (DOTmLPF-P) solutions. For an experimental test article or tactic to have military significance, it must be able to improve how the Navy operates, meet Fleet requirements and define/resolve capability gaps, or explore technological, doctrinal, and organizational concepts desired for operational capabilities for the future. A best practice is to use a project plan to manage overall experiment activities.

A test plan defines exactly how the experiment will be executed. It is very important to develop the test plan collaboratively with the DCAP development discussed in the Analysis Area to ensure alignment of activities. Intellectual properties should be defined clearly and in agreement with the sponsor, as well as the limitations of each. For more information, consult the forthcoming Guidebook to Experiment Analysis,

Key elements for any experiment/study are the Constraints, Limitations, and Assumptions (CLAs), which together bound an experiment.

- Constraints are restrictions or requirements imposed by the study sponsor that define the bounds of the study and restrict the experimentation team's options. Such constraints could look like "Must demonstrate ability to…" or "Operations will only be conducted…"

- Limitations are an inability to fully meet the experiment objectives or fully investigate the issues. A limitation might be “due to the time limitations, only five repetitions of variations in the experiment will be conducted.”
- Assumptions are statements related to the study that are taken as true in the absence of facts. It is critical that these be documented in advance and agreed on by all parties. CLAs can come from any of the functional areas and are included in both the test plan and the DCAP.

3.1.2 Engineering Areas

Engineering areas primarily deal with the installation or modification of systems as well as the removal of the installed system or system modification. Each experiment will have to go through one or more engineering processes to ensure technical rigor and risk management are applied. Different types of platforms (ships, submarines, aircraft, shore) have different process requirements which can result in a complex level of effort. For a more detailed description, consult Section 10.4.

3.1.3 Analysis Areas

Most of the analysis areas are captured in the DCAP. The DCAP is the plan for what data will be collected, how and when it will be collected, and how it will be analyzed. It should contain the primary questions that will be addressed, and how the parts fit together. For a more detailed description, consult Section 8.1. The output of the analysis area is the Test Report – a detailed report about what happened, what it means, and any recommendations. Some experiment teams will also generate a Quick Look Report (QLR). Many CLAs are identified before the experiment. Others are discovered during the experiment. All CLAs and their impacts are included in the Test Report.

3.1.4 Administrative, Manning and Logistics Areas

Ship visits, to ensure optimal communication and ability to meet requirements, ideally are conducted far in advance of installation. Availability will vary and be dependent on ship schedules, but a walk through is crucial to verify installations, locations, and logistics such as power and access. Ship visits also encourage warfighter communication and allows for preparation as well as the opportunity to share first-hand operational knowledge.

To get personnel or equipment on a ship, aircraft, or shore site, there are many complex arrangements and procedures to follow. Naturally, there is a crucial need to recognize and prepare these items to avoid accidents resulted by insufficient sailing preparation. Actions such as Secure for Sea, forms and identification exist to ensure all articles are secured and ready for voyage.

All loose items on deck, as well as cargo, cranes, and anchors must be well stowed and secured. There is also equipment such as motors and hydraulic piping that must be protected to ensure smooth and safe navigation.[6] Just as there is protocol for departure, there is also measures in place for bringing goods and personnel aboard.

When coming on board a Navy ship, there are many important things to keep in mind in terms of both safety and etiquette. For full information on the environment, customs, and living conditions on board a Navy warship, and what is expected while aboard, consult the Ship Rider Orientation Guide.

The ship’s company is the best example to follow when in doubt. Any questions can be directed to the experiment lead. Remember to use common sense and consider fellow shipmates, as upholding a good working relationship with the crew is paramount.

As a guest aboard the ship, preparation and a positive demeanor can go a long way. Display government and activity credentials always (except when underway), especially when boarding the ship. All personnel are subject to the Uniform Code of Military Justice (UCMJ) while on board or at a Naval Facility.

---

[6] Inspection and Securing for Sea (2015)

3.2 Participants

There are three different groups involved in coordinating and conducting experiments. First, the experiment team is comprised of the organizations and personnel that are sponsoring and conducting the experiment. Next, several experimentation organizations exist to mentor and facilitate experimentation teams through the necessary processes. These teams are generally organized to align with the functional areas. Finally, the operational and supporting units that participate in the experiment. The experiment team will be working with the staffs, ships, subs, aircraft squadrons, shore sites, and their Sailors and Civilian to conduct the experiment. Figure 3-2 shows how those functional areas align with experimentation process and some of the activities in each area.

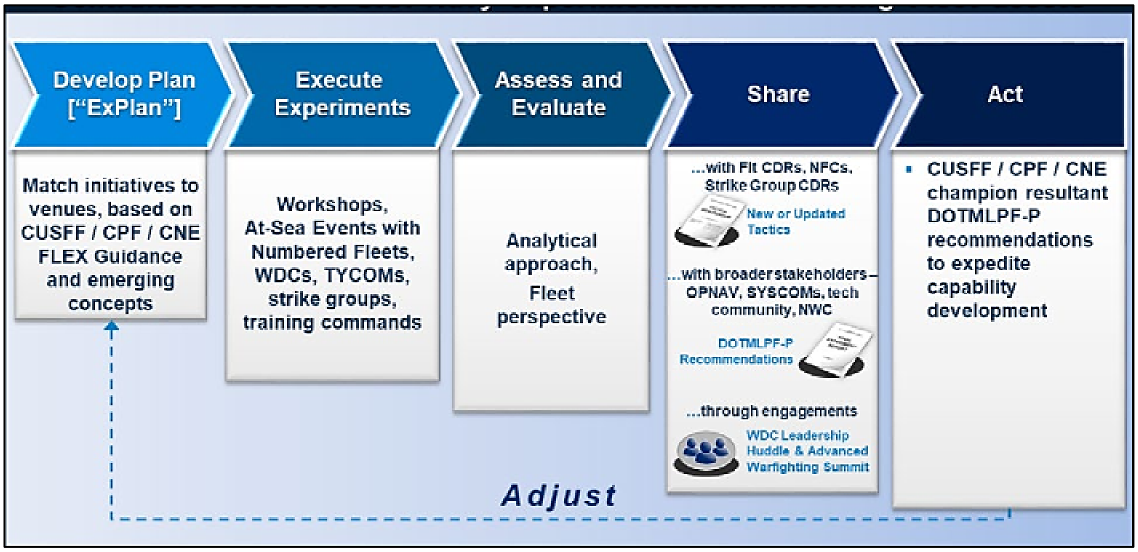

FIGURE 3-2 CENTRALIZED LEAD FOR EXPERIMENTATION INVOLVING FLEET ASSETS[7]

3.3 Process

The experimentation process has four phases: 1) planning, 2) preparation, 3) execution, and 4) analysis and close out. Each experimentation component has different activities that are conducted during each of the four phases. Figure 3-3 provides a high-level view of some of the key activities.

[7] Naval Warfare Development Command

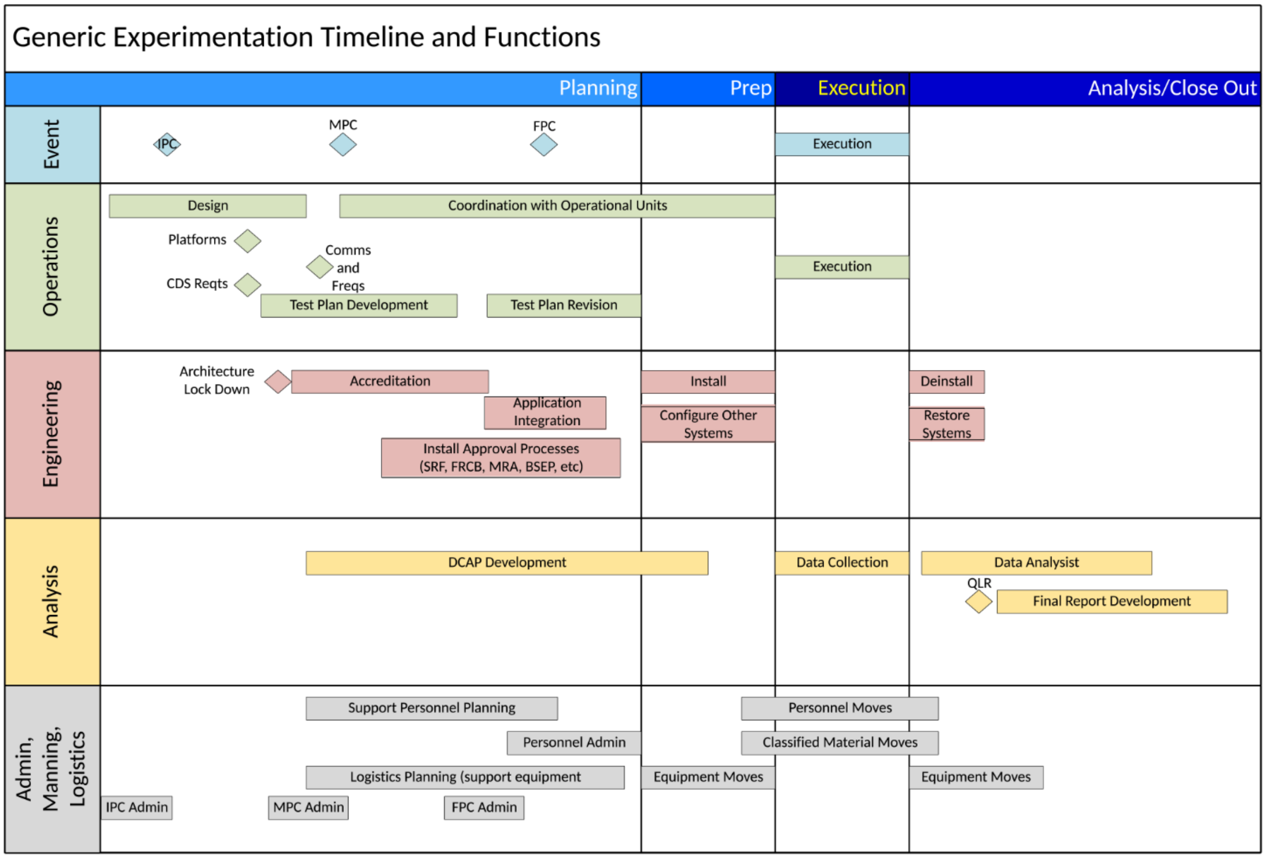

FIGURE 3-3 EXPERIMENT PROCESS AND FUNCTIONS CROSS FUNCTIONAL CHART

Chapters 4 – 7 which follow provide more details about each phase of an experiment process and some of the activities conducted during each.

4 Planning Phase

The planning phase of any experiment addresses the “what” type of experiment the initiative wants to conduct and the objectives to achieve; “when and where” the experiment will be conducted – the event or venue and schedule; “who” is needed to support the experiment – the operational forces, systems, circuits, etc.; and “how” it will be conducted – the DCAP. The planning phase is where all requirements to conduct the experiment are identified and addresses. Often, a project plan or POA&M is generated during this phase.

4.1 Types of Experiments

There are countless routes of experimentation to take when first developing a plan of action. It is important to consider what level of complexity, formality and escalation of effort is appropriate for the hypothesis or idea at hand. For example, is a laser communication test better suited for a tabletop or field experiment? Factors such as risk reduction are essential, as well as reviewing project maturity against desired objectives.

While there are seemingly endless categorizations of experiments, the underlying activities are consistent across the board. A classic experiment is likely the most recognizable type, an event that measures the effect of a change or stimulus. Unless discussing discovery or free-play experiments, most are built around a hypothesis. A clear hypothesis will state the proposed cause and effect relationship, typically featuring an independent (the "if") and dependent (the "then") variable.

A hypothesis may sound like this: **If** lasers are used to communicate in a strike mission – **Then** enemy combatants will be unable to track their location until a strike occurs. While the independent variable is the one manipulated to see how or even if the dependent variable is affected, there are also multiple interfering variables that impact this relationship. The skill level of the sailors, the weather conditions that day, and any other environmental impacts are all considered intervening variables. To ensure understanding of the cause-and-effect relationship, sometimes only one variable is manipulated at a time. However, when there are multiple variables with multiple settings each, to evaluate fuller ranges within a reasonable number of repetitions more than one variable is adjusted at a time. DoD experiments can deviate from the classic experiment formula as they often take on a more informal format. The free-play method is more flexible and can introduce new tools or applications, posing the common question of "what happens if…?" [8]

Below is an overview of the major types of experiments and the activities associated with them; this guide will aim to encompass the activities familiar to these most common types. Regardless of method or activity, all experiments should attempt to uphold the principles of being valid, reliable, credible, and precise.

- *Demonstration experiments* are performance-based and explain a particular scientific concept, such as the classic baking soda volcano experiment. Eddie is interested in volcanic activity and builds a model volcano for the science fair. He already knows the anticipated outcome and there is no question posed nor variables manipulated, the purpose is simply to simulate volcanic activity through a chemical reaction and stage a show for his peers.
- *Hypothesis-testing experiments* start with a clear question or assumption about a subject of interest. A prediction is formed, and the experiment is conducted to render that prediction true or false. A common example would be testing a torpedo system. A research team hypothesizes if a new propellant is used, the effective range will increase. They may also take into consideration the benefits listed below to determine the ultimate combination to maximize results if the hypothesis is proven.

There are several types of activities that experiments can leverage. Examples of common types of experiment activities, from simple to complex, are discussed below. Activities with higher complexity, resources required, and/or operational realism will take a longer time to plan and cost more. For example, live fire events tend to be expensive so instead, validated modeling & simulations (M&S) may be used as surrogates or to better define the 'design' of the live event. Events such as these often require precursor testing (such as medium or lightweight shock testing) in addition to a ship/site visit. These prerequisites will be described more in-depth in future chapters.

- *Workshops* bring together diverse talent to build the structure of experiment as well as generate ideas. Taking on either an informal session or serious deliberation format, it is here that underperforming capabilities are identified and a concept of operations (CONOPS) is developed.
- *Limited Technology Demonstrations are focused on technical performance and are often narrowed to a small set of capabilities.*
    - *Standardized testing using commercial standards,* often in conjunction with military specifications (MIP-SPEC, MIL-PERF, MIL-STD), are often the simplest and best understood type of experiments. Usually performed during the earlier stages of development, they use well defined test procedures and equipment to measure performance against a minimally acceptable value. There is a wide range of applicable test methods that cover items such as material properties, environmental resistance, and weapons performance (i.e., ballistic performance or response to underwater shock events). Often the technology will be required to pass multiple standardized tests before moving forward to larger experimentation efforts. For existing

[8] Department of Defense Experimentation Guidebook, 2019, p. 5

systems, the requirements for standardized tests and performance levels are often defined in the system specification documents.
    - *Laboratory experiments* (often with corresponding M&S efforts) are a logical steppingstone to larger fleet exercises. These are experiments that can be conducted at corporate, university, commercial or government test facilities in a controlled environment to address specific issues. While not as rigid as standardized tests, the test procedures and equipment configurations are usually based on previous experiments. Because of the lack of a rigid, standardized test procedure, there is a strong need to follow the same experimental process of larger experiments in terms of clearly defining the why/what/how of testing and what data will be collected and used to measure the outcome. Testing facilities and methods are often scalable depending on the complexity and maturity of the technology being evaluated. Examples of this would be testing response to underwater shock, in accordance with MIL-STD-901E, using a lightweight shock test machine, a medium weight shock test machine or using a large barge test in a shock pond with explosives.
- *Live Virtual Constructive* (LVC) training is a combination of live and virtual components in a constructive training environment. An example might be a headquarters exercise where the command-and-control systems and methods are tested but all the combat units are simulated. The virtual simulation will see all the parts of the constructive training (whether it be visual or radio messages), just as the live training will see targets and events in its simulated sensors. Naval Integrated LVC Environment (NILE) is one of the events in this category, with the primary focus being integrated kill chains. In addition to virtual elements such as simulation, NILE utilizes test ranges and hardware-in-the-loop tools. These are not the best venues for testing equipment, but a good working computational model can provide a way to assess the impact the new system would have on processes.
- *Field Experiments* contain the greatest amount of realism, as they are conducted with live units using current military technology to mimic conditions they expect to encounter in active operations. Small-scale field experiments sometimes precede large-scale ones if they are not standalone. Multiple trials are conducted and allow experimentation with many solutions. Large-scale experiments provide a full-scale immersive experience, using environmental factors to test technological solutions as well as explorative and safety measures. Since these are a large and often costly affairs, multiple trials are usually not explored in this setting.
- *Wargames* are simulations where new concepts are explored without the dangers of military conflict nor the use of units/ships; mainly in either tabletop or virtual format.

Table 1 has further details about the common activities.

Activity Type	Characteristics	Scale and Frequency
Workshop – Focused Analysis	Exploratory, developmental. Brief, intense effort to discover and generate concepts or examine key experimental issues. Participants identify focus areas and research topics with SMEs. Conducted live but may have virtual components.	Small scale, many each year
Limited Technology Demonstration (LTD)/ M&S	Focused on the technical performance of a technology. An effective LTD should be a methodical test of specific capabilities, looking at single measurements under multiple conditions. May use constructive forces.	Small scale, as required
Seminar-Style Wargame/ TTX	Exploratory. Facilitates discussion forums used to discover and define problem boundaries, pose solutions, and exchange information. Seminars are conducted live but	Small scale, many each year

	may use virtual components. SME and operator participation.	
Systems Wargame, M&S supported	Computer assisted, virtual forces. An event using simulated and emulated systems depicting potential scenarios. Identifies key variables, refines concepts, and assessed alternatives.	Small scale, one or more each year
At-Sea Experiment	Conducted at sea with live forces. Permits participants to see how focused initiatives, technologies or concepts will work in an operational/exercise context.	Small or large scale, varies
Fleet Battle Experiment (FBX)	Integrates smaller efforts as the culminating event. Examines effectivity/validation of CONOPS/TTP when integrated into a higher tactical, operational level of war. Will often involve both live and simulated opponents.	Often large scale, no more than once a year

TABLE 1 EXPERIMENT ACTIVITIES[9]

4.2 Venues & Events

Selecting the place, time, and circumstances - the venue and event - under which the experiment will be conducted, is a key decision during experiment design. Just as there are a multitude of experiments, there are a variety of events ranging in size, scale, complexity, and difficulty to reflect that. When reviewing suitable events and venues, it is important to select the one that best fits the objective, while considering interface and interoperability. For example, a laser in the initial stages of experiment planning may not fit well into a high-TRL event such as Fleet Experimentation (FLEX). However, an early-stage event such as Trident Warrior (TW) may be a good opportunity for the experimentation team to jump into.

Venue is defined as the location and/or activity for conducting an experiment. Properly conducted initiative planning will clearly state experiment requirements as it pertains to units, personnel, and testing environment. Air/water space areas are controlled by Fleet Area Control and Surveillance Facilities (FACSFACs) and have specific requirements to use their areas as per their individual range manuals. Legal and environmental concerns will need to be considered early as well.

Event is defined as a block of time and resources scheduled for the accomplishment of the experimentation, training, tactical development, familiarization, and essentially everything needed to execute and complete the experiment. The event will most likely coincide with an already scheduled Fleet event that an assigned ship/aircraft is already participating in. Training/familiarity requirements must be factored in and added to the plan before support units/personnel get underway for their already scheduled tasking.

Naval experimentation events take place throughout Naval Systems Commands[10], Fleet Commanders Experimentation Programs[11] and ancillary funded venues.[12] The magnitude and focus of these events range from small Limited Technology Demonstrations (LTD) to large scale naval exercises. One example is FLEX, which encompasses exercises, events, and fleet training vignettes to inject capability-based experimentation. It is important to note that participation in these events takes months of preparation and planning. They can consume significant amounts of time and resources; thus, it is important to identify the venues that best fit the need and focus on those instead of taking on too much. Understanding the nuances of experimentation venues is critical for SBIR initiatives in selecting the venue that maximizes the possibility of success.

[9] Experiment Planning Guide (EPG), Naval Warfare Development Command, Revision 2, 2017, p.28
[10] Naval Sea Systems Command (NAVSEA), Naval Air Systems Command (NAVAIR), Naval Information Warfare Systems Command (NIWC) (formerly SPAWAR), Naval Facilities Engineering Command (NAVFAC), Naval Supply Systems Command (NAVSUP), Marine Corps Systems Command (MCSC)
[11] Fleet Experimentation and Fleet Training exercises (Bold Alligator, COMPTUEX, etc.)
[12] Advanced Naval Technology Exercise

This table contains descriptions of the major types of venues/events that are scheduled, as well as several named events that are scheduled on a recurring basis. Events will be distinguished using several criteria: General location, typical TRL level (see 11.3 for definitions of TRLs), format, scale, and entry criteria. This section is not all inclusive of all events, the examples below are just a few of the opportunities available.

	Type	Location	Scale	Format	Time of Year
Advanced Naval Technology Exercise	Event Series	Newport, RI	Large	Hybrid	Spring
Composite Training Unit Exercise	Exercise	Southern CA Jacksonville, FL Cherry Point, NC	Small	Hybrid	6-8 weeks before deployment
Fleet Battle Problem	Exercise	Indo-Pacific	Large	Live	Summer or upon deployment/return
Fleet Experimentation	Program	Various, pending sponsored Fleet	Varies	Hybrid	Various, most events in Summer
Limited Objective Experiment	Experiment	Varies	Small	Virtual	As needed
Large Scale Exercise	Experiment	Varies	Large	Live	Summer
Northern Edge	Experiment	Alaska	Large	Hybrid	Summer
Rim of the Pacific Exercise	Embedded Exercise	San Diego – Hawaii Transit	Large	Live	Late Summer
Technology Innovation Games	Experiment	Varies	Small	Hybrid	As needed
Trident Spectre	Exercise	Fort Story, VA	Small	Virtual	Late Spring
Trident Warrior	Embedded Exercise	San Diego, CA	Medium	Live	Summer
Valiant Shield	Embedded exercise	Pacific Ocean – Guam	Large	Live	Summer

TABLE 2 EVENT INFORMATION

4.2.1 Fleet Experimentation

FLEX is a highly-collaborative, efficient program that aligns initiatives with experimentation events. All FLEX initiatives are screened and briefed by NWDC and their respective review panels, ensuring that high-potential initiatives are authorized to participate in a FLEX event. As shown in Figure 3-2, by combining multiple methods such as workshops, wargames, and at-sea events, FLEX allows for high-visibility and rapid turnaround results using Fleet assets. The Valiant Shield wargames as well as the focused venues of Trident Warrior and Netted Sensors are included in the wide array of exercises conducted under the FLEX umbrella.[13]

In addition to scheduled and recurring events, exercises can also occur on an ad-hoc basis. USNS ships such as Stiletto and Sea Fighter both participate and conduct exercises aboard. For example, Stiletto participated in Trident Warrior and is referenced as a demonstration craft. Sea Fighter serves as a testing ground for new technologies and capabilities. It is important to understand these test ships have much more lenient installation and ship riding restrictions compared to traditional “Grey Hull” ships of the line. Choose appropriate venue and event details from objectives and requirements to prove results are critical in deciding platforms for exception. These discussion and decision processes will be outlined further in forthcoming guidebooks for Installation and Analysis.

[13] Brewer et al, 2018, p. 1

4.2.1.0 Rim of the Pacific Exercise

Rim of the Pacific Exercise (RIMPAC) is the largest global maritime warfare exercise and includes dedicated experimentation opportunities coordinated through FLEX or Trident Warrior. Hosted in Hawaii, this exercise is typically held in the summer of even-numbered years as foreign military forces from the Pacific Rim and beyond travel to participate. The Pacific Fleet Command hosts and administers this biennial event with support from the Coast Guard, Marine Corps, Army, Air Force, and the Hawaii National Guard. To gain an understanding of the magnitude of this event, RIMPAC 2020 concluded in August and included "53 replenishment-at-sea events, 101 pallets of cargo distributed, over 16,000 rounds of small arms munitions shot, over 1,000 large caliber weapons fired, 13 missiles expended, and 1,100 pounds of mail delivered." [14]

4.2.1.1 Trident Warrior

Trident Warrior is a medium-scale venue conducted annually at sea during the summer. Though heavily focused on Information Warfare, this NAVWAR-sponsored event provides a recurring opportunity to work with partners across domains in experimenting with potential initiatives and solutions for identified capability gaps. Although part of the reporting chain within FLEX, it is also fully assimilated into the biennial RIMPAC. This experiment venue allows for exploration into potential initiatives and warfighter feedback early in the process.

4.2.2 Advanced Naval Technology Exercise

Advanced Naval Technology Exercise (ANTX) is a series of individual events (not tied to fleet events) that invites industry, academia, and Government R&D organizations to demonstrate emerging technologies and innovations.[15] For beginner researchers and levels, this event is ideal. Catering to many different capabilities, ANTX provides a lower-risk environment where scientists and engineers can evaluate innovations at the R&D level before militarization to interface at the operational level. Each exercise creates operationally relevant scenarios and environments that are focused on meeting mission priorities and gaps as determined by the Naval Research and Development Establishment (NR&DE) and Fleet Commanders.

In 2015, ANTX began as a demonstration of a single technology at Naval Undersea Warfare Center (NUWC) Newport Rhode Island. In the years since, the events have evolved into multi-warfare center and industry collaborative events that explore and demonstrate hundreds of technologies and tactics in key focus areas. Each year, there is an overall theme to the exercise that focuses on specific gaps and priorities that fall under it. Some examples of these themes include Fight the Naval Force Forward (FNFF 2019) and Naval Integrated in a Contested Environment (NICE 2021).

ANTX/Test-Exercises (ATE) are comprehensive mission-based events hosting various live, virtual, and constructive experiences to support key fleet initiatives. ATE events fall under the ANTX umbrella and have additional focus areas that facilitate the testing and demonstration of technologies and engineering innovations while providing a realistic environment to assess the operational utility of participating technological innovations before potentially becoming militarized and integrated at the operational level. The ATE event conducted in 2019 included projects and technologies focused on joint war at-sea, and long-range surface warfare with an emphasis on distributed lethality.

One of the many events hosted under the ANTX umbrella, the Naval Integration in Contested Environments (NICE) is focused on rapid acquisition, involving a task force hosted by Naval Warfare Centers for the specific purpose of assisting concept development. This is carried out by utilizing a wide variety of industry technology to explore future capabilities. The event is conducted both on site and virtually, providing lower risk in R&D performance as well as the cultivation of relationships with industry partners. This event is held annually around Spring on ranges in Camp Lejeune, North Carolina.

[14] RIMPAC Public Affairs, 2020
[15] Naval Research and Development Establishment. (n.d.).

4.2.3 Composite Training Unit Exercise

The Composite Training Unit Exercise (COMPTUEX or C2X) is a training event, typically month-long, that each USN Strike Group conducts six to eight weeks before deployment. Each ship in the strike group trains in its specialty as the event assembles them to prove they are once again ready to project force and deploy. As an intermediate group exercise with a high TRL level, COMPTUEX is designed to forge together the components of the strike group into a fully operational fighting team. This test event is a crucial part in pre-deployment preparation as well as a prerequisite for the group's Joint Task Force Exercise (JTFX).

4.2.4 Joint Interagency Field Experimentation

The Joint Interagency Field Experimentation (JIFX) program conducts quarterly collaborative experimentation in an operational field environment. The experiments provide an environment where DoD and other organizations can conduct concept experimentation using surrogate systems, demonstrate and evaluate new technologies, and incorporate emerging technologies into their operations. Run by the Naval Postgraduate School typically at Camp Roberts in California, JIFX provides minimal logistical support with not much more than access to a functional airstrip. However, it is an opportunity to bridge the research element of academia (access to faculty) while providing operational evaluations (access to the military student body). Collaboration is seen as an integral element of the program and unless information is closely held (classified or proprietary in nature) it is expected to be shared.

4.2.5 Fleet Battle Problems

Fleet Battle Problems (FBP) were first introduced in 1923 to provide commanders with realistic scenarios that would stretch them to the limits and encourage creative thinking. 21 FBPs were held between 1923 and 1940 and the broad at-sea operations contributed significantly to the success of WWII. Typically, the FBP is a simulated real-world problem that is handed to a senior commander to solve. An alternate senior commander is in turn working against them. The problems are intended to be hard, require creative thinking, and do not have a predetermined winner/loser. They are typically problems that planners have had trouble trying to navigate. At the conclusion of the exercises there is an extensive debriefing where commanders from all levels provide candid feedback on what they saw and learned. The FBP provides an at-sea environment that allows commanders and staff freedom to conduct operations and solve issues in real-time.

This type of exercise has gone unused for some time but was reintroduced to PACFLEET in 2018 by former commander Adm. Scott Swift. What problems future commanders might face is unknown, but it is anticipated that the Fleet Battle Problem in 2021 will prominently feature unmanned vehicles.[16]

4.2.6 Large Scale Exercise

Conceived by the former Chief of Naval Operations, Admiral John Richardson, Large Scale Exercise (LSE) 2021 will be the first major live test examining the fleet's ability to execute Distributed Maritime Operations (DMO), Littoral Operations in a Contested Environment (LOCE) and Expeditionary Advance Base Operations (EABO) in support of the National Defense Strategy (NDS). Designed to execute once every three years, LSE will leverage widely distributed assets across multiple warfighting areas including surface, air, sub, cyber, space, and unmanned systems to assess their cumulative impact, integration, and effectiveness to fulfill mission objectives on the global scale. Participants include multiple fleets and marine expeditionary forces (MEFs). The event will be strategically designed to replicate near-peer adversary warfighting capabilities, tactics, techniques, and procedures. Due to its complex design and high-level maturation, LSE is best suited for well developed, advanced technologies. This venue provides the opportunity to demonstrate interoperability and integration of transitioning SBIR initiatives.

4.2.7 Limited Objective Experiment

A Limited Objective Experiment (LOE) is typically a small-scale exercise that is intended to address a very specific set of objectives or capabilities. For example, a LOE could be conducted strictly as a cost-benefit

[16] LaGrone, 2020

analysis of different camera systems, or to test a new sonar array. They do not require many ships, aircraft, or shore platforms and thus have smaller budgets. The LOE could consist of several smaller experiments that simulate principal warfighter tasks and measure the systems' performance in each. A key advantage of an LOE is that in working on a small number of objectives, efforts can be concentrated more in depth. Refer to Figure 4-1 below for a generic example, while considering LOEs have a smaller range of platforms and supporting resources.

FIGURE 4-1 LOE OPERATIONAL VIEW (OV-1)

4.2.8 Northern Edge

Northern Edge is a fully cohesive exercise providing realistic and inclusive joint training opportunities for Navy, Marine Corps and Air Force units in and around Alaskan land, water, and airspace. It is intended as both a key training event and an opportunity to assess advanced equipment and future operations, putting plans through significant rigor to ensure that what is being proposed is effective and allows service members to hone their skills. Typical objectives include defensive counter-air, close-air support, and air interdiction of maritime targets.

The data gathered during this large-scale event is used to measure the current combat capabilities. Further analysis determines the efficacy of the services integrated and equipment performed in the expansive Alaskan surroundings and simulated Indo-Pacific region. The information is then used in joint publications that are consulted to improve interoperability. Major participating units include U.S. Indo-Pacific Command, U.S. Pacific Air Forces, U.S. Pacific Fleet, Marine Corps Forces Pacific, Air Combat Command, Air Mobility Command, Air Force Materiel Command, U.S. 3rd Fleet, Air National Guard, Air Force Reserve and U.S. Naval Reserve. Exercise Northern Edge normally takes place every other year (on odd-numbered years) around May or June.[17]

4.2.9 Rapid Prototyping, Experimentation and Demonstration

A rapid prototyping, experimentation and demonstration (RPED) project is another ideal route to accelerated acquisition and fielding. This type of exercise is used to expedite the development, exploration and

[17] Russell, 2019

fielding of prototypes. RPED is not exclusive in itself; for example, a RPED team can also gain involvement in an event such as ANTX.

4.2.10 Technology Innovation Games

A Technology Innovation Game (TIG) is a brief exploratory session focusing on collaboration and the development of new concepts. These games can range from wargames, workshops, and demonstrations and gain warfighter insight as well as inform decisions.

4.2.11 Trident Spectre

Trident Spectre (TS) is an annual exercise focused on the integration of operations, intelligence, and technology. Conducted over a two-week period in May, the invitation-only event is organized as a result of continuous planning and collaboration with joint interagency stakeholders. TS is considered the innovation battle lab for Naval Special Warfare and is used to exercise projects that meet USSOCOM operational capability shortfalls and needs from the intelligence community. As an operational experimentation venue, it consists of a deliberate nomination, assessment, and selection of feasible projects in advance.[18]

4.2.12 Valiant Shield

As one of the largest U.S. wargames hosted in the Pacific Ocean, Exercise Valiant Shield often involves a carrier strike group and other combatant vessels. This event is an LSE and typically weeklong. The exercise takes place in the late summer, every other year (on even-numbered years) following the RIMPAC exercise. While RIMPAC is meant to be more inclusive for all partners and allies, Valiant Shield is U.S.-only and typically covers more sophisticated warfighting tactics and systems. The primary focus of this event is inter-branch cooperation and monitoring of at-sea units.

4.3 Framework and Timeline

As shown in Figure 4-2 Step B, each experiment requires the development of an analytical framework to ensure the required technical rigor is applied to develop the experiment design and execute data analysis. The framework is often developed by the analysis functional area during the planning phase. From the framework, a timeline or schedule of events for execution is developed and defined in a test plan, developed by the operations functional area. The complexity and underlying variables must be identified and strategically integrated into the experiment design. The framework is also defined in the DCAP.

The backbone for conducting an experiment is supported by the military problem statement and top-level objective(s). Establishing client/stakeholder relationships is critical to verify/validate objectives and sub-objectives not only align with the problem statement but are also achievable given the experiment venue and available resources.

[18] Trident Spectre 2021 Information Guide

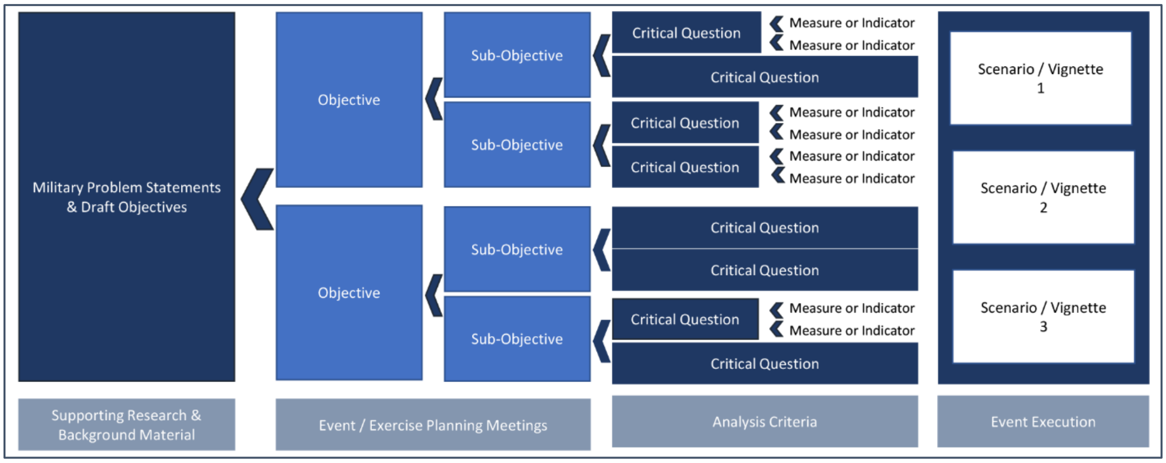

FIGURE 4-2 ANALYTICAL FRAMEWORK EXAMPLE[19]

The experiment process and timeline (defined in the test plan) can be thought of as like the process of building a house. Homes cannot be built at the last minute; there are processes to follow for it to turn out well. Just as each house has a unique address and qualities, each experiment is unique by its very nature of trying to learn something new. While each instance is unique, they require the following of fundamental steps and planning to be successful. The planning timeline is dependent upon stakeholder availability, venue, required resources, and priority. These details will be revealed over time and must be carefully integrated into the plan.

Like many, this process can be broken down in to three sections: before, during and after. In each section, it is crucial to remember and focus on what is important. For the house, there is everything that goes into the preparation (blueprint, ordering supplies, scheduling workers), then the build out (where the plan is executed), and then the building is ready for use. An experiment needs to be planned, executed, and then analyzed with the data that is collected.

With a house, a box of nails is just one store run away, but those granite counter tops need to be custom cut and that takes time. While some things for an experiment can be handled on the spot, most issues need to be taken care of well in advance (60 days, 90 days, or even a year). It requires a significant amount of effort beforehand for an experiment to go smoothly, and without a coordinated plan, things can go bad quickly. There are many different primer models used to shape planning, as modeling usually begins with the objective. What is the core of the central idea being worked? This is the OV-1 model and shows the high-level, conceptual architecture. Refer to Figure 4-3 for an example.

[19] Navy Warfare Development Command

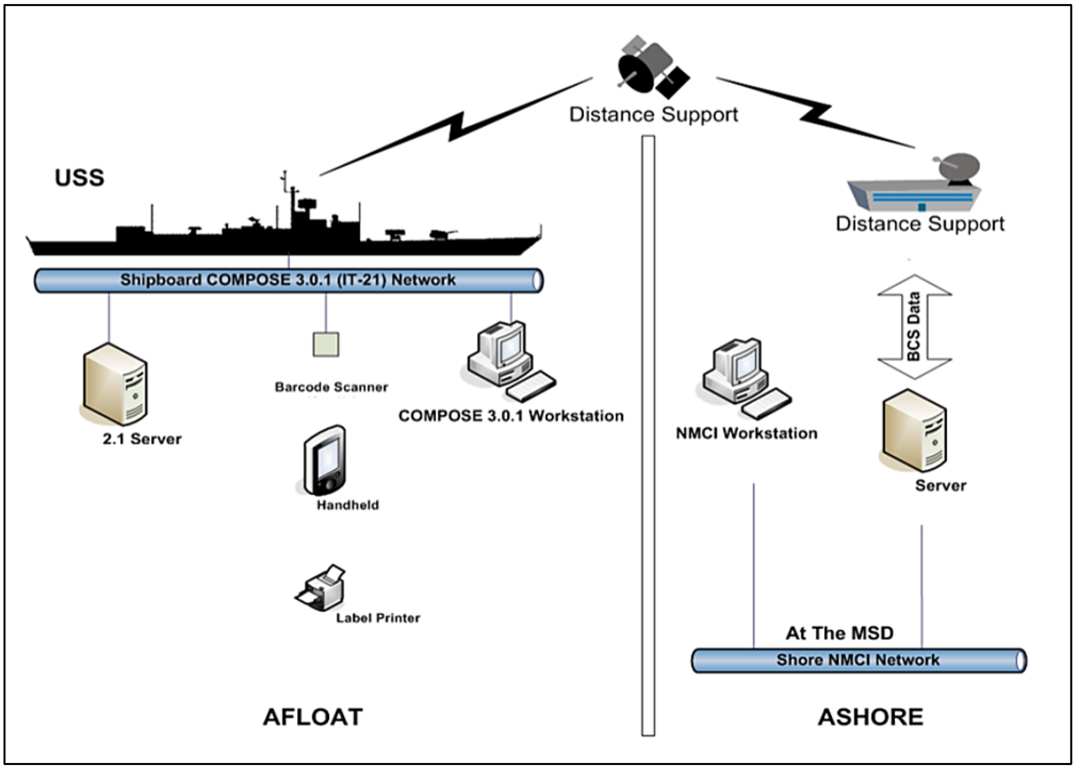

FIGURE 4-3 OV-1 EXAMPLE

Most exercises have several planning conferences leading up to the main event, and experiment planners are expected to be ready to answer key questions at these conferences. These meetings occur months in advance and serve as milestones that can help keep an experiment on track. In most cases, they will have specific deliverables that will be used to enable the most important work of the conference— coordination. During the conference, details concerning the scenario can be developed, personnel can be assigned to positions, requirements and dependencies paired up with support elements and the Master Scenario Event List (MSEL) refined. If any of these elements negatively impacts the experiment, it is critical that the event organizers and/or the experiment sponsor needs to be made aware. They should also be made aware if the experiment can support or enhance an already-scheduled exercise element.

Typically, exercises have three planning conferences: Initial, Main, and Final. The requirements and timelines can vary from venue to venue but in most cases, experiments are expected to have the problem formulation developed before the initial planning conference begins. Some elements may be requested beforehand to get invited to the planning conference. A basic problem formulation will include: the problem statement, primary objectives, proposed solutions, and primary measures. By the main planning conference, a fully developed experiment design with critical questions should be in hand along with a draft DCAP. By the final planning conference, all documentation for the experiment should be finalized and ready to go. Most planning conferences are spaced out between one- and six-month separations to enable participants to complete tasks and prepare for the next milestone. This can mean that an initial planning conference can occur a year or more before the actual event.

4.4 Formulation

Formulation is a foundational step for all experiments and is a lot like the start of a major research paper. The main questions are established as well as scoping so that there are recognized boundaries of how far the research will go and what topics it will and will not deal with. A good formulation will answer the question

"Why does this experiment need to be conducted," it will have a basic plan, and will help shape the team should any specialties be required. The problem statement must be thoroughly developed before any background research can be conducted. The problem statement should clearly address the whole issue and not just focus on the hypothesis, as the purpose is to identify gaps and capabilities. The more structured and refined a problem statement is, the easier it will be to form a specialized team and proceed with the experiment.

4.4.1 Problem Statement and Objective

The problem statement is a core element of the formulation and establishes the main purpose. A problem statement in the military will commonly address a capability gap, need, condition or obstacle, and how that affects the mission. To develop this statement, it is best to research history to avoid recreating the wheel or relearning past lessons.[20] An accompanying list of objectives helps describe hopeful achievements, lessons to be learned, and important attributes. Shown as Step B in Figure 4-4, this measure provides the pathway for data collection, experiment design, reporting structure, and additional engineering/operation development that may be required.

4.4.2 Entry Criteria and Matrices

An entry criterion is used to determine the start date of a given test activity. It also signals the introduction of the next "level" in the process when the test design or execution is ready to start. Depending on the experiment's installation type, several engineering processes exist with documents (such as test plans) that must be presented for approval. These processes culminate with the Risk Assessment Request Message which will be sent to obtain TYCOM or FLTCDR authorization for the FLEX activity.[21] An experimental matrix is a table containing all the trials to be conducted, with the number of trials and the levels assigned to each factor within them included. An effects matrix is used to calculate the impacts or results.

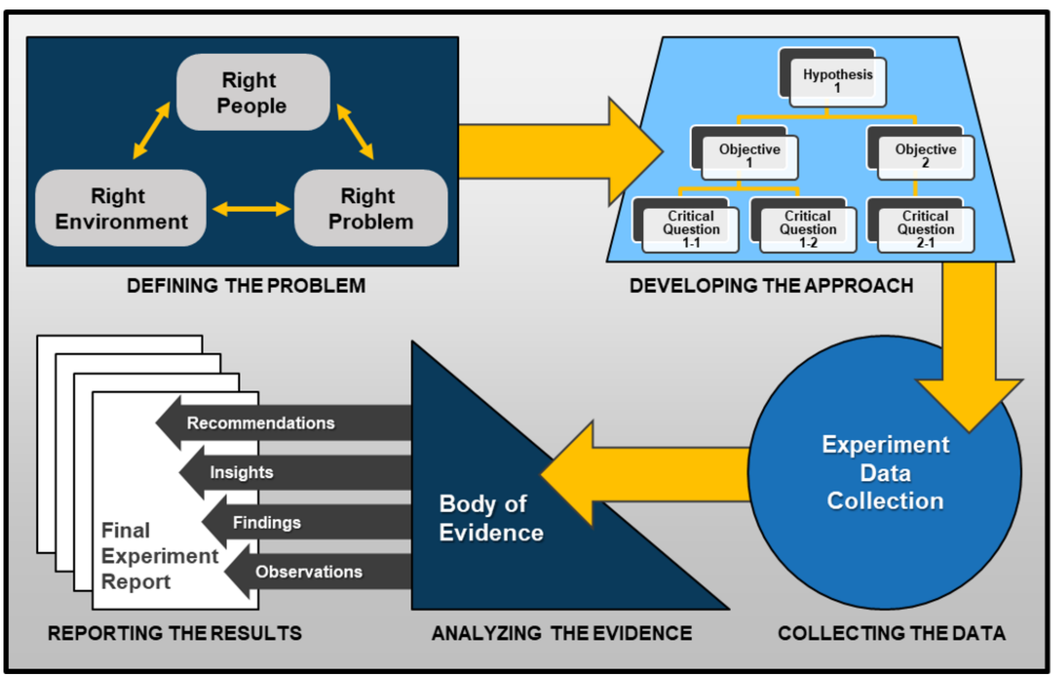

FIGURE 4-4 ANALYTICAL STEPS IN EXPERIMENTATION[22]

[20] "A Design for Maintaining Maritime Superiority", Chief of Naval Operations, 2018
[21] Appendix Q, Navy Modernization Program Management and Operations Manual, 2019, pp. Q-71- Q-86
[22] Navy Warfare Development Command

After establishing the problem statements and key questions, the formulation process can conclude, and the follow-on stages can begin. This piece of the process may often take up the bulk of the time in an experiment design process, with months of development culminating in the major experimental event (which may only take a few days, followed by a month or two of analysis and reporting). There are several major planning areas to explore.

4.4.3 Operational Planning & Development

Operational planning is often dependent on the size of the exercise, as the deliverables increase along with the size/scale of the event. For smaller and ad hoc events, the process may be streamlined. The tasks and deliverables commonly associated with operational planning include: Primary Model designs, POAMs, operational sequence diagrams, event overview briefs, public affairs overview papers, event execution manning plans, and coordination for Hotwash/Quicklook. Joint Capability Areas (JCAs) are a major DoD framework that facilitates operational planning and development. JCAs are grouped into nine separate areas and assist collaboration by providing a universal lexicon and realistic solutions. By incorporating JCAs into the operational planning process, the gap between user and experimenter is further bridged and is considered a great tool in experimental development. For more details, consult Section 11.2, Appendix B.

4.4.4 Event Execution Design Plans

A typical design plan will contain details on all elements of the experiment, answering the questions *"What is going to be done? How will it be done? How will it be measured? And what will be done with the results?"* Plans need to be in place for the handling of any sensitive or classified information. The event execution design plan will likely contain information on the following areas:

1. Overarching event design documentation.
2. Detailed experiment objectives.
3. System engineering plan(s).
4. The integration of systems and interfaces over the replicated operational network.
5. Coordination with focus and system leads to identify and mitigate technical risk reduction by identifying and executing the specific technical tests to be performed that address the risk matrix.
6. Details on all test events and ensure all test events are properly scheduled, staffed, and executed.
7. Site visits as necessary to verify existing network architectures at experiment sites.
8. Verified operations in accordance with specifications and assist in troubleshooting and identifying methods for rectification.
9. Security and the transportation/storage of large or sensitive systems.
10. Safety and environmental issues for both pre-and post-testing.

4.4.5 Experimental Variables

Revisiting the torpedo example from Section 3.1, the gap identified (Step 1) could be: "Due to improvements in enemy detection capabilities, submarines need to be able to launch torpedoes from longer ranges." In studying the problem (Step 2) the researcher could study the cost for the seekers available, how much the detection capabilities have improved and what allies are doing. The objective (Step 3) is to have a better torpedo, but "better" needs to be defined. Does that mean more range, better guidance, faster travel times? It is important to be specific. In identifying the factors (Step 4), the controllable variables need to be isolated, including the type of launch platform, the propellant used, the support systems, and the size and type of the target. An uncontrollable factor would be the environmental conditions such as sea states or water temperature. There can be many factors to account for, from very large factors like the fire control system used, all the way down to participants involved and their skill levels. The *levels* or settings of each factor in this example would be the number of torpedoes tested, the seekers, or warhead size. *Response* is the output of the experiment. In this case, how long did it take for the seeker to lock on to the target? Was it able to maintain lock while the target was moving? Did it ever lose lock and need to re-acquire? Based on the warhead size and the target capabilities, what is the expected kill rate (or probability)? These and possibly many others could be measurable outcomes that may have been influenced by the factors and the levels that were modified. Figure 4-5 below gives a visual look at the relationship between factors, levels and then response using the torpedo experiment as an example.

Factors (variables, inputs)	Levels (settings)
Seeker Type	Acoustic/Magnetic/Combo
Warhead Size	Small/Med/Large
Propellent	Type A/B/C
Guidance Package	Old/New

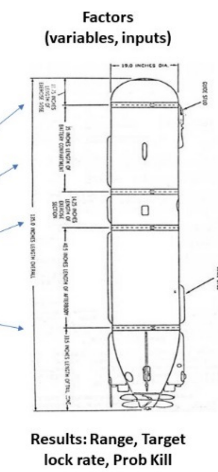

FIGURE 4-5 VARIABLE RELATIONSHIPS

With four factors and multiple levels each, a researcher might not want to test every combination, after all, launching a significant number of torpedoes can get expensive very quickly. Consulting a SME might enable the researcher to eliminate combinations based on the SMEs years of experience and from previous tests. For example, if propellant B requires significantly more volume than A or C, it might not be possible to pair it with a large warhead (due to the minimum required range and the volume limitations). Through research, a best set of combinations can be selected (Step 5).

4.4.6 Engineering Planning

Planning will encompass a set of early requirements to enable installation on Navy ships, aircraft, or facilities for experimentation purposes. Because the experiment may have some impact on systems, it is important to understand and mitigate the risk to the operational units. With that in mind, testing and certifications are often required before the experiment can be executed on a Navy ship, shore facility, aircraft, or submarine. Most of the time, the sponsor and/or participating acquisition manager (PARM), or its assignee, will coordinate the testing and certification activities. Keep in mind that not all requirements and deliverables will be needed in every situation. There are minimal requirements for simple experiments. However, there are a larger set of requirements for complex experiments that interface with ship systems. Gaining an awareness of what may be involved can help the experiment owner understand the process better and why it appears to be a long timeline. Additionally, this may generate a greater understanding for the formulation of questions about how the process relates to a particular experiment. For further details and information about these processes, consult Section 10.4, Appendix D.

4.4.7 Analysis Planning

Analysis planning and assimilation should be started as early as possible, the time invested helps to focus the team and create a strong foundation. The analytical techniques intended to apply will dictate the data that needs to be collected and the precision required. Building a detailed plan and following it are essential in providing the credibility and validity areas of results. Preparatory study, experimental design, and reliable assessment all contribute to achieving analytical rigor. During the planning phase, personnel in the analysis functional area develop the analytical framework previously discussed and shown in Figure 4-4 Step B and in Figure 4-2. The analytical framework is the basis for identifying the data that needs to be collected to achieve the experiment objectives, as well as under what operational conditions (the vignettes) it will be collected. The when, where, and how the data needs to be collected is codified in the Data Collection and Analysis Plan (DCAP). The framework also informs the operational area of the vignettes needed to develop the test plan. The test plan defines the step-by-step process for executing the experiment.

4.4.7.0 DOTmLPF-P Criteria

It is critical to understand the implications of new technology, and its potential impact across multiple DoN programs. DOTmLPF-P are major areas typically associated with non-materiel recommendations. It is critical that recommendations align with new and developing technologies to enable greatest positive impact. A system's technical capabilities, intended use, and user interface will impact future integration and applicability across multiple DoN programs. While a system might be designed to address a specific warfighting area, it can impact numerous focus areas throughout the DoN. When conducting capability-based assessments, each of these should be examined to assess the potential impact. For example, a new communication method could have the following impact:

Organization – Better coms might enable a new tactical formation.
Training – Do units need to be retrained on the systems, is it simple or complex?
Materiel – Any procurement issues with the new system need to be addressed, how will it interface with existing systems?
Leadership – How will command and control be impacted?
Personnel – Will additional personnel be needed, or will it be automated and enable a smaller crew?
Facilities – Does the new system require a higher classification and locations need to be secured?
Policy – Are there going to be any issues sharing this tech with international partners?

By looking at questions like these one can evaluate the impact the new technology will have, not just in the primary area that it will be used but also the second and third order effects.

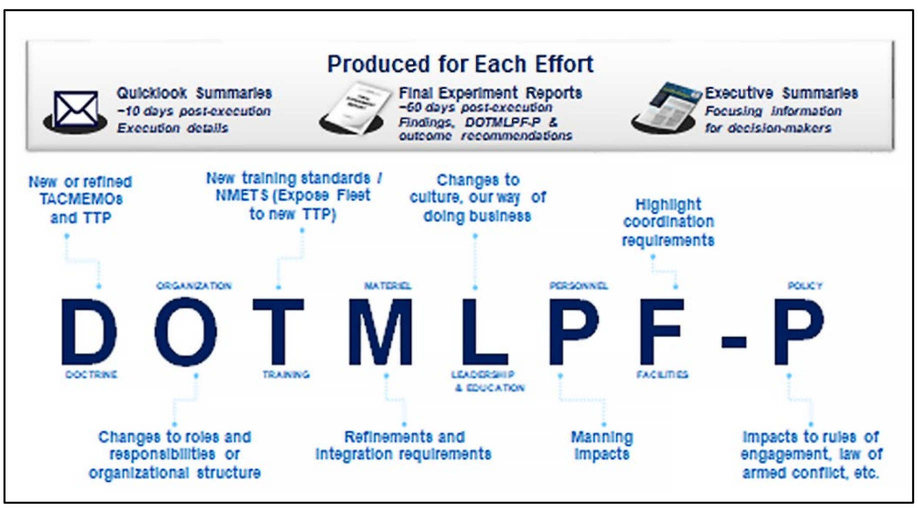

FIGURE 4-6 FLEX PRODUCTS[23]

5 Preparation Phase

The preparation phase is the time where the experiment's operational and technical environment is established, the team and materials are moved into place to conduct or execute the experiment, and the operational units configure their systems and position themselves to support.

[23] Navy Warfare Development Command

Early in the preparation phase, the heaviest focus is on the installation of experimental systems and data collection equipment. The engineering functional area coordinates and oversees these activities. The administrative functional area coordinates the embarkation of the experimentation team, classified material, and other logistical requirements.

During and after the systems are installed and the experiment team embarked, the ships and submarines position in their designate locations, all operational units will configure supporting systems, establish communications paths and networks, and conduct all other activities needed to provide the required operational environment.

Some preparations can begin even before a venue is selected, and then refined once a venue is selected. Because the preparation is a liquid plan, additional needs and sometimes setbacks can arise, requiring the team to modify the plan. Securing funding, strengthening the hypothesis/problem statement, and forming strategies and schedules all occur during this stage. This phase will culminate to the conducting stage, which can run anywhere from days to weeks. The approach to performing these events can take many forms as discussed in Table 1 earlier.

6 Execution Phase

The final phase is the execution/collection and requires conducting of the experiment. The crew and/or staff of the operational units work with the experimentation team to conduct the experiment per the experiment plan and collect data per the DCAP. Important factors/settings involved are recorded and the outcomes are measured. These results are shared and may be retested many more times with the goal of providing enough information for future researchers and warfighters alike. Given the costs and consequences of live fire testing, only a small number of replications will likely be conducted. As such, it will be important to collect data at every stage of the kill chain to isolate as many cause and effect relationships as possible. Execution through completion require processing the data and publishing the results, perhaps in a journal of some sort, or the data might be deemed classified and can only be shared through appropriate channels.

While most experiments execute as planned, the operational environment can be dynamic. The operational units, SBIR Experimentation Team, and the Experiment Team work collaboratively to address changes to schedule and operational environments and define the impact of any changes to the experiment. The engineering functional area addresses unforeseen technical problems and provides advice and recommendations for addressing any deviations to the experiment plan. The analysis functional area has the critical role of ensuring all the required data is being captured as well as documenting any CLAs and deviations.

Some experiments require members of the experimentation team or equipment to move between operational units during the experiment. These personnel moves could be caused by an equipment casualty or the need to embark a particular SME on different units at different times to achieve the experiments objectives. The administrative/manning/logistics function handles these moves.

7 Analysis and Close Out Phase

Once an experiment is conducted and the data collected, the functional areas oversee the restoration of systems to their pre-experiment condition, removal of the experimental systems, disembarkation of personnel, and transfer of data collected during the experiment. The final step is to examine and interpret the data and see what can be learned from it. The DCAP should contain details for the original data plans, not just what will be collected, but also what will be done with it. It is critical to enlist the aid of SMEs to help interpret the results, leading to the validation (or invalidation) of the hypothesis. While the DCAP is important, it is encouraged to also pursue paths outside of the plan. The data may contain additional information that was not anticipated and should be checked for potential insights.

7.1 Results

The results of the analysis and the experiment itself are captured in a test report. Occasionally, a QLR is also generated. It is often perceived that for an experiment to be judged a success, it must validate the hypothesis. According to the DoD Experimentation Guide, this view has propagated a risk-averse culture that in turn has limited the utility of some experiments. Rather than pushing the boundaries of a topic, the “safer” route is selected. To counter this culture, it is necessary to institutionalize a new understanding of what experiment “success” and “failure” really mean and to focus on the true intent for any experiment—to advance knowledge. Ultimately an experiment needs to be considered a success if it can establish a clear cause and effect relationship between the key variables. A successful experiment is not whether the hypothesis is validated, but instead the expansion of knowledge. If more research is required, there is the option to conduct the experiment again (if funding allows) or terminate the effort. Fortunately, a well-planned experiment can result in at least one of the following impacts: new or updated models, updated requirements for new experiments, solution changes (or the filtering of failed ones), operational transition, new TTP development, transition to rapid fielding, and/or integration into programs both existing and new.

8 Summary and Closing

Innovation and experimentation in the defense sector have one key attribute, that of serving the Fleet. The ability to rapidly deliver the latest technological advances that could support or solve urgent warfighter gaps is highly valuable for defeating near-pier adversaries in an ever-changing world. By capturing publications, processes, and best practices, the goal is that experimentation managers can more effectively assess risk, associate/allocate funding, and greatly increase the possibility of carrying out successful technical experiments. Though the experimentation process may appear daunting, it is important to remember that options are available and there are multiple avenues to helpful resources through the SBIR program. It is important to not let fear or intimidation hamper the exploration of possibilities and assistance available to innovators. Exciting methodologies specially tailored for SBIR communities and the venues they support remain to be documented and codified as part of this ever-evolving process. By exploring groundbreaking Naval concepts and fostering invention, innovators will continue to be agents for change.

9 Acronym List

#
2 Kilo
3-M Maintenance Action Form
2M Miniature/Microminiature
3M Maintenance and Material Management
3-M Maintenance and Material Management
3-MC Maintenance and Material Management Coordinator
3-MPR 3-M Performance Rate

A
ACR Alteration Completion Report AIT Alteration Installation Teams
AIRS Airworthiness Issue Resolution System
ANTX Advanced Naval Technology Exercise
AP Acquisition Plan
AS Acquisition Strategy
ASN(RDA) Assistant Secretary of the Navy (Research, Development and Acquisition)
ATE ANTX Test Exercises

B
B Billion
BAA Broad Agency Announcement

C
C4I Control, Communications, Computer, and Intelligence
C5ISR Command, Control, Communications, Computers, Combat Systems, Intelligence, Surveillance and Reconnaissance
CBA Cost Benefit Analysis
CCB Configuration Control Board
CFR Code of Federal Regulations
CLA Constraints, Limitations, and Assumptions
CM Change Manager
COMPTUEX Composite Training Unit Exercise
CONOPS Concept of Operations

CPAF Cost Plus Award Fee
CPAT Cost Plus Award Term
CPIF Cost Plus Incentive Fee
CSP Comprehensive small business Subcontracting Plan
CUI Controlled Unclassified Information

D
DCAP Data Collection and Analysis Plan
DCR DOTMLPF Change Request
DEVGRU Development Group
DFARS Defense FAR Supplement
DMO Distributed Maritime Operations
DoD Department of Defense
DoDI Department of Defense Instruction
DoN Department of Navy
DoN-SEC Department of the Navy SBIR Experimentation Cell
DOTmLPF-P Doctrine, Organization, Training, materiel, Leadership and Education, Personnel, Facilities and Policy
DPM Deputy Program Manager

E
E3 Electromagnetic Environmental Effects
EABO Expeditionary Advance Base Operations

F
FACSFAC Fleet Area Control and Surveillance Facilities
FAR Federal Acquisition Regulation
FBP Fleet Battle Problems
FFF Form Fit Function
FFRDC Federally Funded Research and Development Center
FIMS FLEX Information Management System
FLEX Fleet Landing Exercises
FNFF Fight the Naval Force Forward
FRC Fleet Readiness Center
FY Fiscal Year

G
GRSL Groupsail

H
HCA Head of Contracting Activity
HITL Human-in-the-Loop

I
IAW In Accordance With
IPL Integrated Priority Lists
ICPL Integrated Prioritized Capability List
IDEF Integration Definition
IFC Interim Flight Clearance

ILS Integrated Logistics Support
IP Intellectual Property
IPL Integrated Priority Lists

J
J&A Justification and Approval
JCA Joint Capability Areas
JCIDS Joint Capabilities Integration and Development System
JFCOM Joint Forces Command
JIFX Joint Interagency Field Experimentation
JTFX Joint Task Force Exercise

K
KO Contracting Officer

L
LAR Liaison Action Request
LCC Life Cycle Cost
LCM Life Cycle Manager
LCS Littoral Combat Ships
LOA Light-Off Assessment
LOCE Littoral Operations in a Contested Environment
LOE Limited Objective Experiment
LOI Line of Inquiry
LSE Large Scale Experiments
LSRB Laser Safety Review Board
LTD Limited Technology Demonstrations

M
M Million
M&S Modeling and Simulation
MCSC Marine Corps Systems Command
MDD Material Development Decision
MEFs Marine Expeditionary Forces
MIL-PERF Military Performance Specification
MIL-SPEC Military Specification
MIL-STD Military Standard
MoE Measures of Effect
MoP Measures of Performance
MRA Mission Readiness Assessment
MSC Military Sealift Command
MSEL Master Scenario Event List

N
NATIP Naval Aviation Technical Information Product
NATOPS Naval Air Training and Operating Procedures Standardization
NAVAIR Naval Air Systems Command
NAVFAC Naval Facilities Engineering Command
NAVAIRINST NAVAIR Instruction

NAVSEA Naval Sea Systems Command
NAVSUP Naval Supply Systems Command
NAVWAR Naval Information Warfare Command
NAWC Naval Air Warfare Center
NDA Non-Disclosure Agreement
NDAA National Defense Authorization Act
NDE-NM Navy Data Environment – Navy Modernization
NDS National Defense Strategy
NICE Naval Integration in Contested Environments
NIWC Naval Information Warfare Systems Command
NLLIS Navy Lessons Learned Information System
NMP-MOM Navy Modernization Process Management and Operations Manual
NNMSB Non-Nuclear Munitions Safety Board
NTIRA/SMART Navy Tool for Interoperability and Risk Assessment/Submarine Modernization and Alteration Requirements Tool
NSA Naval Supervisory Authority
NWDC Navy Warfare Development Command

O
O&M Operations and Management
OPNAV Office of the Chief of Naval Operations
OPNAVINST OPNAV Instruction
OPSEC Operations Security
OPTASK Operational Tasking Orders
OSD Office of the Secretary of Defense
OSHA Occupational Safety and Health Administration
OSIC On-Site Installation Coordinator
OV Operational View
OWLD Obligating Work Limiting Date

P
PAO Public Aircraft Operation
P&E Prototypes and Experiments
PD Policy Directive
PEO Program Executive Office
PFC Permanent Flight Clearance
PICO Pre-Installation Check Out
P.L. Public Law
PLOA Probability of Loss of Aircraft
PM Program Manager
PMAP Protective Measures Assessment Protocol
POAM Plan of Action & Milestones
POC Point of Contact
PoR Program of Record
PY Planning Yard

Q
QA Quality Assurance
QLR Quick Look Report

QMS Quality Management System

R
R/R&D Research/Research and Development
RDT&E Research, Development, Technology and Engineering
RFF/RFS Request For Forces/Support
RFP Request for Proposal
RIMPAC Rim of the Pacific Exercise
RMMCO Regional Maintenance and Material Coordination Office
RPED Rapid Prototyping, Experimentation and Demonstration

S
S&T Science & Technology
SBA Small Business Administration
SBC Small Business Concern
SBIR Small Business Innovation Research
SBP Small Business Program
SC Ship Change
SCD Ship Change Document
SCN Ship Conversion Navy
SDCP Shock Deficiency Correction Plan
SDM Ship Design Manager
SE Early Systems Engineering
SFAF Standard Frequency Action Format
SIL Systems Integration Lab
SME Subject Matter Expert
SOF Special Operations Forces
SOVT Systems Operational Verification Testing
SOW Statement of Work
SPM Ship Program Manager
SRF Service Request Form
STTR Small Business Technology Transfer
SUPSHIP Supervisor of Shipbuilding
SV Systems View
SYSCOM Systems Command

T
TAES Technical Area Experts
TAT Technical Assessment Team
TDP Technical Data Package
T&E Test and Evaluation
TE Technical Experimentation
TEMPALT Temporary Alterations
TIG Technology Innovation Games
TNTE2 Tactics and Technology Exploration and Experimentation
TPOC Technical Points of Contact
TRL Technology Readiness Level
TSCE Total Ship Computing Environment
TTA Technology Transition Agreement

TTP Tactics, Techniques, and Procedures
TV Technical Standards View
TW Trident Warrior
TWH Technical Warrant Holder
TYCOM Type Command

U
UAV Unmanned Aerial Vehicle
UAS Unmanned Aircraft System
UCMJ Uniform Code of Military Justice
USN US Navy
USSOCOM United States Special Operations Command
USV Unmanned Surface Vehicle
UUV Underwater Unmanned Vehicle

V
VCOC Venture Capital Operating Company

W
WG Working Group
WSESRB Weapon System Explosive Safety Review Board

X
XO Executive Officer
X-RIC Pseudo-Repairable Identification

10 References

1. Alberts, D. S., &; Hayes, R. E. (2002). *Code of Best Practice for Experimentation*. Washington, D.C.: DoD Command and Control Research Program. Retrieved from http://dodccrp.org/files/Alberts_Experimentation.pdf
2. Brewer et al. *Tomahawk Fleet Experimentation, Systems Engineering, and Delivering Capability to the Fleet*. Johns Hopkins APL Technical Digest, 2018, www.jhuapl.edu/Content/techdigest/pdf/V34-N03/34-03-Druce.pdf
3. Chief of Naval Operations, *A Design for Maintaining Maritime Superiority*, Revision 2. December 2018, https://media.defense.gov/2020/May/18/2002301999/-1/-1/1/DESIGN_2.0.PDF
4. CYLITIX. (2020). Trident Spectre 2021. Retrieved January 07, 2021, from https://www.cylitix.com/homepage/event-directory/ts21/
5. Department of Defense. (2010, December 1). Electromagnetic Environmental Effects Requirements for Systems (MIL-STD-464C). https://www.dau.edu/cop/e3/DAU%20Sponsored%20Documents/MIL-STD-464C.pdf
6. Department of Defense Experimentation Guidebook, Office of the Under Secretary of Defense for Research and Engineering, 2019.
7. Department of Defense Guidebook for Electromagnetic Spectrum Survivability, (July 2020) Washington, DC: Office of the Joint Staff J-8.
8. Department of Defense Prototyping Guidebook, Office of the Under Secretary of Defense for Research and Engineering, 2019.

9. Experiment Planning Guide (EPG), Naval Warfare Development Command (NWDC), Revision 2, 1 July 2017.
10. Joint Fleet Maintenance Manual (JFMM) COMUSFLTFORCOMINST 4790.3, Department of the Navy, Revision D, Volumes V-VI, 16 October 2019.
11. LaGrone, S. (2020, September 08). *PACFLEET Preparing for 2021 Unmanned Fleet Battle Problem*. Retrieved January 06, 2021, from https://news.usni.org/2020/09/08/pacfleet-preparing-for-2021-unmanned-fleet-battle-problem
12. Legal Information Institute. (n.d.). 10 U.S. Code § 2366b - Major defense acquisition programs: Certification required before Milestone B approval. https://www.law.cornell.edu/uscode/text/10/2366b
13. Management and Operations Manual Appendix H: Alterations to Ships Accomplished by Alteration Installation Teams, Navy Modernization Process, 2015.
14. Management and Operations Manual Appendix Q: Fleet Experimentation and Technology Demonstration, Navy Modernization Process, 2019.
15. Metzger, D. (2019, July). What Are SBIR Data Rights and Why Are They Important? Retrieved March 01, 2021, from https://www.sbir.gov/sites/all/themes/sbir/dawnbreaker/img/documents/Course9-Tutorial2.pdf
16. NAVAIR Airworthiness and Cybersafe Process Manual, NAVAIRINST 13034.1F, 13 APR 2016.
17. Naval Research and Development Establishment (NR&DE). (n.d.). Advanced Naval Technology Exercise (ANTX). Retrieved January 06, 2021, from https://www.secnav.navy.mil/agility/Pages/ANTX.aspx
18. Navy Warfare Development Command. (n.d.). Fleet Experimentation. Retrieved from https://nwdc.navy.smil.mil/FIMS
19. Office of the Director, Defense Research and Engineering. (July 2009). Technology Readiness Assessment (TRA) Deskbook. Department of Defense. Retrieved from https://www.acqnotes.com/Attachments/Technology%20Readiness%20Assessment%20Deskbook.pdf
20. Program Executive Office. (2017, February 14). PMW 160 Tactical Networks Program Office. https://www.public.navy.mil/navwar/PEOC4IandSpace/Documents/TearSheets/PMW160_FactSheet_2017_DistroA.pdf
21. RIMPAC Public Affairs. (2020, August 31). *Multinational Exercise Rim of the Pacific 2020 Concludes*. Commander, U.S. Pacific Fleet. https://www.cpf.navy.mil/news.aspx/130716
22. Russell, C. (2019, May 30). *Exercise Northern Edge 2019 comes to a close*. Retrieved January 07, 2021, from https://www.af.mil/News/Article-Display/Article/1861223/exercise-northern-edge-2019-comes-to-a-close/
23. Secretary of Defense. (2014, November 15). *The Defense Innovation Initiative*. https://archive.defense.gov/pubs/OSD013411-14.pdf
24. Secretary of the Navy. (2019, 26 March). Defense Acquisition System and Joint Capabilities Integration and Development System Implementation. https://www.secnav.navy.mil/rda/
25. Ships Business. (2015). *Inspection and Securing for Sea*. Retrieved January 15, 2021, from http://shipsbusiness.com/procedures-for-securing-for-sea.html

11 Appendices

The following sections contain supplemental data, tables, and figures to help further navigate through experimental development.

11.1 Appendix A – Preparation

As explored in earlier chapters, the preparatory piece of the process may often take up the bulk of the time in an experiment design process. This process is dependent on asking the question of what one wants to learn. The key to a well-prepared design plan is to gather knowledge and perform research on the subject to be studied. Once a research question is formed, the experiment team will define the variables required to study it. During this process, the research team considers the possible variables, both those that can be controlled and those that cannot be controlled but must be monitored. The goal is to understand what variables impact the answers to the questions.

11.1.1 Experiment Design

Experiment design is simply the creation of a procedure set to test the hypothesis. Refer to Figure 11-1 below for an overview of the process and steps entailed to verify the question posed. Experimentation using Fleet assets has complex challenges, as the process involves intensive coordination and planning, with internal and external dependencies. The plan needs to have contingencies incase changes need to take place. An unexpected deployment could cause a unit to drop out of an exercise; it is important to anticipate the impact if a primary asset were not available and is there a way to mitigate it. Experiment design in this arena, especially on naval ranges or in nationally controlled air/water space are processes that require approval from federal organizations. Specialized action officers and doctrinal physical domains dictate requirements for specific approvals and determine the level of effort required to operate assets and technologies legally. The consequences of an ill-planned operation can be astronomical; therefore, safety is a large factor as well.

FIGURE 11-1 EXPERIMENTAL DESIGN CYCLE

11.1.2 The Inputs

How a knowledge gap is identified is outside the scope of this document, but the gap itself is a critical element to the overall design as the source for the questions that need to be answered. It is critical to know exactly what the gap is; understanding what the warfighter needs will in turn dictate the questions that need answering, which leads to the experiment itself for answers.

11.1.3 Define/study the Problem

"Ask the right questions if you're going to find the right answers." - Vanessa Redgrave. To define a problem, one must ask a lot of questions, so ask them. What exactly is the gap? Can it be filled, either completely or partially? Can the team deliver even more than needed? Is the need clearly communicated and understood? Keep asking questions. Consider what resources (such as SMEs, funds, personnel) are available to help with this. Ask about the timeline for a solution (perhaps a 90% solution now is better than a 100% solution in three years). Establish the CLAs and ensure that the stakeholders agree to them.

11.1.4 Determine the Objective

Having defined the problem statement, the research team can start to solve it by establishing the objective for the experiment. Is there a need to know more and the experiment is to see what happens if X, or is there a hypothesis to test ("if A is done, B is expected"), or is the goal to demonstrate that X can do Y? There may be more than one, and there may be sub-objectives.

11.1.5 Establish the Variables / Factors

For every system there are inputs and outputs. The inputs for the experiment are factors that can be controlled through selections. These are called independent variables. The outputs or results are the dependent variables. Assuming the hypothesis is correct, their values are dependent and will change based on the selections for independent variables. In an experiment, an example of an independent variable could be the speed the ship is traveling (with permissions, this can be controlled and adjusted). An uncontrollable variable might be the weather or sea state; the weather cannot be controlled, but the researcher can record the weather conditions to determine impact. The dependent variable may be the accuracy of detecting a target, or time required to do so. Many of the dependent variables will become, or will be closely associated, with the Measures of Performance (MoPs) and Measures of Effect (MoEs). In experimental design literature, variables are often referred to as factors.

While not controlled, some uncontrollable factors can be minimally managed. If weather is predicted to have an impact on the experiment, it is possible to request that the replications of the experiment not be conducted back-to-back, but instead be spaced out over the duration of the exercise. The weather of each replication is uncontrollable, but by spacing them, the likelihood of having varied weather conditions is increased. A similar method could be used for sea states and lighting conditions.

11.1.6 Security & Classification

Critical technology is technology essential to the design, development, production, operation, application, or maintenance of an article or service that makes a significant contribution to the military. This technology can include design and manufacturing know-how, technical data, software, keystone equipment, and inspection/test equipment.

All classified material must fit the following requirements: proper marking, never be discussed in public places and instead in secure areas with approved equipment, stored in an approved storage container and always must be under the control of an authorized person. Even if a document is unclassified, it may be marked as Controlled Unclassified Information (CUI). CUI is still data that needs protection as required by laws, regulations, and/or government wide policies and should be marked per DoD instruction. In addition, consider the Distribution Statement of the document, from Distribution A (approved for public release) down to the more stringent Distribution F (further dissemination only as directed by controlling office or higher DoD authority).

11.1.6.0 Data Rights

The government is under an obligation not to disclose SBIR Data for a period known as the "Data Rights protection period." A SBIR/STTR policy directive came out in 2019 that changed the protection period for SBIR/STTR funding agreements to a 20-year period and eliminated the previous extension or "roll-over" provision. So, from the date of award and for 20 years thereafter, the government may not disclose technical data of the SBIRs that are participating in Phase I, II, or III funding. (It is important to note that a Phase III award is separate funding.) Additionally, the government cannot compete technologies that contain SBIR data as it could be disclosed in solicitations.

To constitute SBIR data, the following attributes must be in place:

1) It is recorded information.
2) The information is technical in nature.
3) The information was generated under a SBIR/STTR funding agreement and has been marked appropriately.

In accordance with the SBIR Data Rights, the Government must enter into an appropriate non-disclosure agreement (NDA) with any non-Governmental entity that is authorized to receive SBIR Data (that is subject to SBIR Data Rights) during the protection period, except as otherwise permitted.[24] SBIR Data that is recorded can be source code, drawings, sketches, equations and formulas, reports, technology descriptions, final reports, and any other recorded information of any of these types. Protections, however, would ONLY apply to concepts or ideas if they are in writing. The protections only apply to technical data, not to any cost or pricing information. To qualify as "technical data" in this context, the data must be related to a SBIR technology being developed in Phases I-III and is something that the SBIR company wants to protect from general disclosure, as opposed to the type of data it presents on its website. Any proprietary data developed by the SBIR company from its own funds does not fall under this protection.

The new policy directive also requires that data falling under this protection be clearly marked and if not marked appropriately or not marked within six months, it will not qualify for the protection. This applies civilian and military agencies. The new data rights clause is very different from the previous clauses and so it is important for the SBIR company to carefully review the funding agreements before signing them. Special language needs to be included that acknowledges and affirms the new data rights clause. It is important that the SBIR Data is marked properly. If a PowerPoint presentation is made to the government, it must be marked with the SBIR Data Rights legend; and everyone in the room should be verified to be with the government upon presentation.[25]

11.1.7 Design the Experiment

Having established the variables or factors, the next step is to set the levels and determine the combinations that will be tested. At each step, it is important to look back at the main objectives and use them to frame the next. If the objective was to establish the ability to accomplish a task while traveling at sea, then the ship speed will not need to be controlled, just recorded. On the other hand, if the objective was that it could be done while traveling within the speed range of 5-20 knots, then it may need to be a controlled variable, with levels set for 5, 10, 15 and 20 knots, so that the task can be tested at each level.

A full factor experiment would require testing every combination of factors and levels. If there are several factors and lots of levels, this could quickly become an impractically large number of test iterations. In such cases, a "smart" combination of sets needs to be selected so that the impact each variable has on the system can be determined.

[24] SBIR and STTR Program Policy Directive, 2019
[25] Metzger, 2019

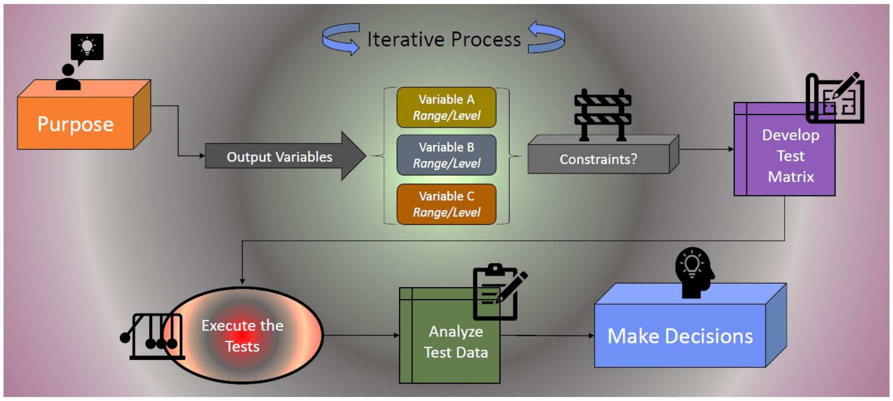

FIGURE 11-2 DESIGN PROCESS AND MEASURES

11.1.8 Execute / Collect the data

"*No plan survives first contact with the enemy.*" - Moltke the Elder. This could be the easiest step; the plan is in place and just needs to be followed. But what if things go wrong? It is important to plan for contingencies. What if a test does not go the way it was expected to? Is there enough time or resources to repeat the experiment? What if a resource believed to be available and thus planned for is not, what is the back-up plan? What if there is an unexpected "visitor", would the presence of a foreign element add to or prevent the experiment? It is also imperative to collect as much data as possible/practical. Some elements may not have been part of the original plan, but if recording an additional bit of data is relatively easy and inexpensive, then do so. That data might be useful later and can prevent the unfortunate position of having to repeat an experiment because something was not recorded and later discovered to be important.

11.1.9 Analyze the Data

Once the data is in hand, it is ready to be processed. From this point, the data needs to be checked and cleaned for any anomalies, and basic summary statistics conducted. Organizing it into tables, or even creating something as simple as a histogram might provide insight on the makeup of the data and the relationships between the variables. Building graphs and creating other visuals can help see what the data is saying. General exploring can be useful, but the focus should be on following the analysis plan that was created and determining the expected results/relationships are supported.

The purpose for analysis is to sift through the data and take it from its raw data form to information to knowledge and hopefully gain some insights along the way. Analysis is about finding the significance or implications of the experimental findings. It is about determining the relationship between variables, understanding the amount of impact they have on each other and potentially recommending areas for future research. It is important to relate results to the aims of the experiment through summarization and explanation and to lead cross-initiative analysis of the generated data. Oversight of specific initiative analyses for internal consistency and connection can contribute to overall program goals. Thus, it is important to maintain documentation, previous event results, and associated relevant data.

When a study is conducted at the end of the experiment, a bunch of numbers remain: the input values (or settings) and the output values (or results). The challenge is to extract from the data a meaningful summary of the behavior observed and a meaningful conclusion regarding the influence of the experimental treatment (independent variable) on participant behavior. Statistics can provide an objective approach to performing this process.

This document will not discuss the various ways statistics can be used or the statistical tests that can be used to process data. At this time, it is simply stated that there is more to data that just the mean and the variance. One key pitfall that must be avoided is thinking that correlation equals causation. Correlation is when two sets of data seem be closely associated with each other. An example of the pitfall would be the link between ice cream sales and sunburn. A quick look at that data would show that the values are highly correlated. Every month that has high ice cream sales has many sunburn cases and months with low ice cream sales do not. The correlation says they must be linked; therefore, ice cream causes sunburn. This is of course not true, it just so happens that during summer months when it is hot, people tend to like ice cream and going outside. Summer/hot weather is the cause for both, and there is no real connection between ice cream and sunburn. This type of issue is why it is critical to have SME availability during the analysis phase to ensure that the interpretation of the results is appropriate.

11.1.10 Interpret Results

One of the most important things to look at when interpreting the results are the objectives; and it is also important to ask the questions again that needed answers. Now that the data is available, can the questions be answered? At the start of the experiment there were one or more hypotheses proposed (see section 11.2.4). Does the data support them? Going into the experiment assumptions were made regarding cause-and-effect relationships. Were the relationships observed?

11.1.11 Output Reports

The final stage of an experiment/campaign is reporting the results in collaboration with stakeholders and Focus Area Leads. This reporting can come in many forms, from a simple bullet background paper that lists the very basics of the experiment and the results, all the way up to a formal paper that will be published in a technical journal or as a standalone book.

The real goal of a study is not the data, but the answers to the key questions. To truly do that, the results must be published in some manner in a report, a presentation, or some other way of disseminating the findings. The key stakeholders will want to know not just the conclusions, but the work involved. There may be a very specific format required (such as a report related to an acquisition process for submittal to the Requirements Oversight Committee (ROC)); or the format may be more open in structure. It is crucial that clarity and transparency be maintained. If the results of the experiment were ambiguous then the report should reflect that. If there is solid evidence of a finding, that should also be clearly communicated.

The format for the simpler reports varies considerably and depends heavily on the audience. The key to any good report is clarity. It needs, in the simplest way, to share the results of the experiment and what was learned. It might need some background on the problem and the basics of the experiments structure, but the real focus should be on the findings.

One common way to share experimental results is a scientific paper. Most scientific papers have the same structure in the writing and publishing process. In general, there are seven sections: Title, Abstract (a short summary of the paper), Introduction (provides background information and includes the hypothesis), Materials and Methods (the details about how the experiment was done), Results (the relevant data collected from the experiment), Discussion/Conclusion (explains the data and how it either supports or does not support the hypothesis), and Literature Cited (lists references relevant to the experiment).

As mentioned above, the most important is the results section. This is the heart the research paper. Here, statistical analyses of the collected data are presented using text, tables, and figures. Remember, statistical analyses do not prove anything, they only provide guidelines as to the reliability and validity of the results.

11.2 Appendix B - Data Collection and Analysis Plan

“If you fail to plan, you are planning to fail.” - Benjamin Franklin

Properly designed and executed experiments enable the advancement of knowledge, understanding of systems/operational concepts within the operating environments and situations of relevance. These processes can be compared to that of building a house. There are three main phases to the construction of a house, design, building and finally, the move in. An experiment is similar as it is planned, executed, and then the data obtained is analyzed.

There are many steps that must take place prior to groundbreaking, such as consulting with the buyer to ensure requirements and expectations are established, followed by drafting of blueprints. Developing stakeholder relationships paired with detailed planning are just as essential to ensure the experiment design and analytical strategy align with experiment objectives/sub-objectives. From this point, available resources and potential alternatives can be investigated. Just as a house is more likely to succeed if the preliminary steps are done, experiments entail an iterative process whose level of success is dependent upon many sub-elements and planning intricacies. The planning process is a continuous cycle that must be monitored throughout experiment design and execution.

As the supporting structure for experimentation, the DCAP provides analytical rigor specific to the experiment objective(s) and associated line(s) of inquiry (LOI). Developing a predetermined research method is a critical strategy to observe, collect, assess, and report on experiment data. Research design and methods are different, but closely related, because good research design ensures that the data obtained will more effectively answer the research question. Each experiment must have a DCAP that is tailored to the experiment that entails a holistic approach towards data collection and analysis procedures.

The DCAP comprises the main elements of the experiment, including the problem statement, objectives/sub-objectives, LOIs, CLAs, critical questions, measures of performance (MOPs) and measures of effectiveness (MOEs). Furthermore, it provides the analytical rigor required to collect and assess quantitative and qualitative data. Just like a house, the DCAP could be thought of as the blueprint, providing detail and clarity to see the individual parts/components, as well as enable participants to see the big picture. Each step in the experiment will add a piece of knowledge helping to fill the requirements, just as each brick laid helps establish the house. With each experiment, a body of knowledge advances, starting with the knowledge already in hand and expanding on it when possible and reinforcing where necessary. Developing a research method is a critical strategy to observe, collect, and report on data obtained during an experiment. Research design and methods are different, but closely related, because good research design ensures that the data obtained will more effectively answer the research question. For any type of experiment, the DCAP explains the end-to-end structure.

The study purpose is typically the first thing to be established, answering the question of why conduct the study. If it is associated with a capability or knowledge gap, that will be prominently featured as the problem statement. If the purpose for the experiment is more demonstration in nature, there might not be a problem statement, but instead have a key feature that is to be highlighted. Next, the hypothesis is developed, asking the question of what cause-and-effect relationship is trying to be established. This will in turn lead to the metrics and data that is required to answer the questions and establish the cause/effect relationships. Each of these elements will be discussed in detail below.

11.2.1 Problem Statement

"If I had an hour to solve a problem, I'd spend 55 minutes thinking about the problem and five minutes thinking about solutions." - Albert Einstein

One of the most important parts of a journey is knowing the destination. One can have great snacks, a car full of gas, and the best playlist, but without knowing the destination it is hard to map the route. It is the same for an experiment campaign, starting with the end goals in mind is critical to overall success. While it would be wise to not take Einstein too literally and spend 95% of the time on this step, it does merit significant time, thought, and energy to ensure the problem is fully understood before undertaking the design. A big part of that understanding is background research, often called a literature review. The experiment should **add** to the body of knowledge. If something has already been done, it is not adding, only repeating (some things bear repeating to provide verification that the previous endeavor was accurate). In general, however, experiments should generate new information. This background research is the foundation the house will be built on.

A clearly articulated problem statement should address three areas: The capability gap, the key stakeholders, and the needed capability. Important sources that can be used for researching these are Integrated Prioritized Capability Lists (ICPLs), Integrated Priority Lists (IPLs), Navy Lessons Learned Information System (NLLIS) and Joint Lessons Learned Information System (JLLIS). ICPLs and IPLs capture the major warfighter gaps and provide details on why they are important and who the stakeholders are. NLLIS and JLLIS provide tactical and operational lessons learned for experiment planners and fleet operators. These lessons can provide the framework for development of doctrine, TTP, concepts of operations, or for improving naval and joint operations of current combat systems, including systems approaching initial operational capability.

It is important to periodically review the campaign plan from top to bottom to verify that everything is still applicable, as well as to incorporate any new information. The results of one experiment might indicate that the original goal is either unreachable or too easy. An experiment might have gone even better than expected and provided data that makes the next planned experiment unnecessary. When this happens, the experiment can be restructured to look at another factor. In some cases, data might have been lost and the experiment needs to be repeated. It is also possible that the original problem has changed, and what was once a capability gap is no longer a problem. Perhaps a new gap has been found, or an unexpected result might indicate that the tactic/technology can solve a need that was not part of the original scope. The big picture might be completely different than it was at the start of the project. For this reason, periodic review of the problem statement can ensure that it is still applicable to the problem at hand.

11.2.2 Research

"Those who do not learn history are doomed to repeat it." - George Santayana

Research is a key element to any experiment. Literary research starts with a general idea and a need to know more, much like an informal discovery experiment. It is often not hypothesis-central, instead the focus is to see what happens in the environment in which this idea is tested. In most cases, the intent of an experiment is to add to the existing body of knowledge, and/or to make a comparison of the new system against the current standard. The answers to questions examining the current standard are critical and will be used not only in the pre-experiment development phase, but also in the post experiment analysis phase. If it is known how the old system was tested, the new can be tested in the same way to do a side-by-side comparison. Without knowing the answers to these questions, it is difficult to demonstrate that the new system is better.

11.2.3 Objectives

Many activities are done "just for the fun of it" such as playing a game of tag, jumping out of an airplane, or climbing a mountain. And that may be true for a small number of experiments, they are done just to see what happens. It may be that A, B & C have never been done at the same time, but now doing that may make a difference. However, a good experiment, especially for an experiment campaign, the experiment needs to have an objective or goal in mind. Also, at each step within, the experiment needs to have sub-objectives that lead back to the main objective.

The problem statement and the objectives are not the same things, but they are joined at the hip. Being able to clearly state the problem helps to define the objectives. One of the easiest ways to flush out an initial list of objectives is to look at the problem statement and ask critical questions such as: "*Why is that important?*" This chain of questions and answers helps flush out the objectives and leads to clues as to how they can be measured.

Once the main objective is established and drafted, it can be broken down into parts. Building up the list of objectives and sub-objectives is a matter of asking the how, what, and why. And then ask again until arriving at a specific task or identifying a metric (see metrics in Section 11.2.5). For example, if the overall objective is: "*I want to sell my widget/idea*" or "*I need to impress the decision maker*", how can that be accomplished? If the widget or idea is proven to work, what makes it better than the other options? How is it unique? How can that be demonstrated? Keep asking questions because the answers to these questions will create a list of potential objectives. Not every combination of question and answer will need to be a documented objective but going through the process will help to make sure that a complete picture is created. In theory, completing all the sub-objectives will in turn complete the objective above it. If this is not the case, then consider what else needs to be done to achieve that.

On a multiday road trip, each day might have a destination as a sub-objective and the snacks might be the topic for a separate set of objectives. Example objectives might be healthy and limited snacks, so dinner is not ruined. Just one change could impact multiple objectives. Lots of snacks are needed to replace lunch to avoid a stop and make better time. Perhaps an accident heard on the radio causes a change in the overall route.

In that case, many of the sub-objectives might also need to be changed. The traveler may receive news that a friend needs help, in which case the main objective is changed and head out in a completely different direction. Similarly, changes might need to be made to an experiment campaign plan as it is being executed. New knowledge leads to adjustments; a new gap might be discovered, or the gap originally targeted might already be filled.

11.2.4 Hypothesis

A hypothesis is a statement or claim that has yet to be supported with data. It proposes a cause-and-effect relationship between two elements of concern. A well-crafted hypothesis helps to focus an experiment and points it in the right direction for what and where to investigate. It typically has two parts, the independent variable in the "if" half and the dependent variable in the "then" half. Examples could be as simple as using a new fuel additive to extend the range the fleet can cover between refills, an approach of changing factor A to improve factor B. It is important for the planners to document their hypothesis before designing the experiment to ensure that the correct data can be collected.

A hypothesis test is the process of determining if there is enough evidence support the proposed cause/effect relationship. However, statistics can be a little tricky in a way, statistical evidence can never prove that something is true. Instead, evidence is used to establish that the likelihood an opposite statement (also called the alternative hypothesis) is so small that the alternative must be false and in turn the original statement can be "considered" true. In a court of law, the public wants to know if someone is innocent or not, however a defendant cannot be found innocent. Instead, the opposite question is asked to prove/disprove the question of guilt. If there is not enough evidence, they are not found innocent but found **not** guilty. The assumption is first that a person is innocent and may prove they are not. In a similar way, the data will not prove the original hypothesis, it proves that the alternative is not true.

Consider testing a new sensor to establish that the new sensor is better at identifying targets then the legacy system. The hypothesis could be as simple as "the new sensor is better." But "better" is very vague and is hard to define. Is the new system more cost effective? Is it more compact? Does it have a higher ID rage? Is that under specific conditions? A higher quality hypothesis might be, "By incorporating the new sensor package, target identification processing time can be increased 50% by eliminating the need for second looks." This version states not only the "if" (incorporating the new sensor package) and "then" (increase target identification processing time 50%), but also adds a possible reason for the impact. Now there are two things to test for: the processing time, and the number of times a second look was required.

Consider an example where the primary hypothesis is that a new sensor has a higher detection rate than the legacy system. The alternative hypothesis is that the two sensors are not different, but in fact produce similar results. If the experiment provided the following data:

	Day light conditions	Low/No light conditions
Legacy	75%	55%
New system	80%	75%

These are different, but are they different enough to reject the alternative hypothesis that the two systems are equal in capability? In effect, is there enough evidence to convict? In this case (unless the values represent a very large sample set) for daylight hours there is not enough evidence to dismiss/reject the alternative hypothesis, and the two systems might in fact be the same. As a result, there is no data to reject the alternative and thus cannot say anything about the original hypothesis. For the low/no light, the likelihood of getting values this different while they are the same is very low. Thus, the alternative hypothesis can be rejected while supporting the original hypothesis that the new sensor does have a higher detection rate under low/no light conditions.

11.2.5 Metrics

One of the main questions that the DCAP needs to answer is what is going to be measured or collected. There are many ways to describe a metric, but the most important way is simply good versus bad. A good metric will help answer the most important question, "**So what**?" What was learned? Does it make a difference? Was an important point discovered? Depending on the topic being studied, different questions will be important; and in turn will influence the metrics that need to be collected. It is important to not focus on one category of data as being superior to another. Like most things dealing with analysis, what is important depends on many factors. The best advice regarding metrics is to think about the "so what" and do not ignore data that can be harvested. If in doubt, it is best to over-collect and sort it out/analyze it later.

Good metrics have three qualities: Valid, Reliable, and Credible. *Valid* metrics means that the measurements being taken are true indicators of the situation. Looking at the color of a strip of bacon would be a valid way to measure the thoroughness of cooking; however, in the case of a roast beef, the color only reveals the surface. The temperature at the core of a roast determines if it is cooked to completion. In this situation, one would ask "What is a valid metric in relation to the subject being measured and what is being determined?" *Reliable* metrics are consistent metrics. In this case, one would conduct an experiment under the same circumstance as a previous experiment and then see if the same result be recorded. With the thermometer and roast, a reliable thermometer inserted into different areas of the same roast should have identical readings. This can be used for metrics that are qualitative in nature (see Quantitative versus Qualitative below) or are opinion-based. *Credible* metrics are synonymous with what can be believed or trusted. Asking a colorblind person a question that is dependent on reading a digital number makes sense, but asking that person if a certain shirt clashes with pants would yield a non-credible opinion.

In general, metrics can be sorted three different ways: Quantitative versus Qualitative, MOE versus MOP, and level of measurement. In many cases, the way a question is asked can impact the way it is measured and the type of data that is to be collected. So, think about the question and be sure the "so what" can be answered. To help explain the differences in the data types consider an example of testing a new pistol below and look at the types of measurements that might be made and how they fall into each category.

11.2.5.0 Qualitative vs. Quantitative

Some things can be counted or measured and given numerical values, others are hard to determine numerically. Those metrics that can be counted are considered "quantitative," and while they might not always be the easiest to count and measure, the value that is recorded is not subjective nor a matter of opinion. How many times did a thing happen? How far can it go? How fast is it? Questions that fall into the "qualitative" category are subjective in nature and are opinion-based, such as "Do you like it?" and "How much do you like it?" It is important to rank these qualitative options and list advantages and disadvantages.

For the pistol example, one might ask the following: *How much does it weigh? How many rounds can it hold?* These are measurable or countable and as such are quantitative. Next, one might ask the following: *Does it feel comfortable in the hand? Does it look good with this outdoor gear?* Those considerations are a matter of opinion. Depending on the hand-size and the shape of available grips, a comfort and a preference scale can be used but the results are very subjective. Some questions can be both qualitative and quantitative depending on how the topic is addressed: *How much recoil is there? Is it too much? Does it impact accuracy?* In the hands of a trained shooter the recoil might not be significant, but for a novice shooter the same recoil could be significant. The recoil can be counted by setting it on a stand and firing the weapon, but to measure the impact would be less accurate perhaps? It can also be tempting to try to turn a qualitative observation into a quantitative one, by rating the comfort level 1 to 5. The answer will be a number, but treating those subjective opinions like true quantitative metrics can be dangerous.

11.2.5.1 Measures of Performance vs Measures of Effect

MOPs and MOEs are often mistakenly thought to be the same, but they are as different as the questions they answer. MOP's answer the question "what?", whereas MOE's answer the question "why?" MOPs are all about **performance**, they often point inward and are focused on actions and what was done. They answer questions along the lines of: “What did you do? What can you do?” With the pistol example, one might ask about its effective range or misfire rate. Did the warhead explode? MOE’s are about **effect**, or impact. They are typically focused on the second; and even the third order of effects from actions. For MOEs the questions are: *“Why did you do it? What was the impact? Did you see the behavior you wanted? Was the building targeted sufficiently damaged?”* For the pistol example, an MOP might be as follows: *“How many rounds per minute can be fired? What is the number of rounds capacity? What is the muzzle velocity?”* The MOEs would be stopping power, ability to penetrate soft skin vehicles, or comfort level in-hand.

Some people look at MOPs as answering the question, “Are we doing things right?” Likewise, they are looking at MOE’s as answering the question, “Are we doing the right things?” Below is an example involving an experiment to study the impact of going to the gym, losing weight, and making new friends. In this example, free gym memberships are sent to collect metrics (MOP’s and MOE’s).

The MOP’s and MOE’s for this example can be quite simple. An MOP might be related to the tasks of going to the gym and completing a workout. The potential MOEs are the impact targeted with this new behavior, body shaping, weight loss, etc. The following questions are asked: “Did you go to the gym three times a week? How many hours of cardio did you complete in the week? Did you interact with other people while there?” These generate the MOP’s. Some of the answers are very easy to count with a yes or no answer. Some answers are harder to measure, such as greeting the employee checking IDs only, or making eye contact without verbal communication.

In theory, it is good to define what is required to count a MOP as successful, but not always best to let the subject being measured know this in advance. The question “How many times did you go to the gym this week?” is better than “Did you go three times?” The second question will only receive a yes/no answer and provide limited opportunity for detailed analysis. For instance, a person attending three times a week could look equal to a person attending six times a week. The first question will get better data, and if three is a key number, the detailed data can be used to split results into two groups and generate the data received from question two. Additionally, question two could create a bias in the study. The way the question is asked could impact the subject’s behavior, and in turn, the study. If the subject hears a question about three times a week, the subject may see that as a goal. The subject would then stop at three or not bother with a second trip. This is an example of why one should carefully consider how a question is asked.

The MOEs can be straight forward: “Are you losing the weight that you set out to lose?” “Do you feel more self-confident about your looks?” “Are you making new friends that will encourage healthy activities?” As with the MOPs, some are easy to measure, like the weight. Step on a scale every day for a month and record the value, what does the needle say? Making new friends can be a little harder to measure. The subject may say, “I see the same people there every time I go, so I recognize them but never talk to or see them outside the gym” versus “I talk to them all the time both inside and outside the gym, we are practically best friends now.” But where on that scale is a “Yes” and where is a “No?” It is critical to understand the difference between the two and how they work together to provide a complete picture. There are four combinations of ways that MOPs and MOE’s can interact listed in the table below. When looking at the combinations, it is good to remember that the purpose for an experiment is to gain knowledge about something not previously explored in this setting.

MOEs / MOPs	Accomplished	Failed
Accomplished	Good (#1)	It depends (#3)
Failed	It depends (#2)	Bad (#4)

Combination #1 looks like it is all good news, a big win for the team. But remember that the purpose for the experiment is to gain knowledge. Always ask “What we can learn from this?” and “Why does that matter?” First the MOPs indicate what can be done and concerning the MOE’s, #1 provides evidence that the assumptions on the association between dependent and independent factors seems to be true. By doing the things set out to do, it is possible to achieve the intended effects. But besides the fact that those things are possible, what was learned? Having the right metrics enables the analysis and gaining understanding of the "so what." How did the things done impact the effects targeted? How are they linked?

As good as #1 looks, #4 looks bad, but does not have to be considered a complete loss. Remember the purpose for an experiment and think about what was learned. If tasks were not completed, why? If the desired results were not produced, why? The cause-and-effect relationship seems to be intact, but is it? Is the model of the situation sound? What can be learned from the data? Even a "failed" experiment can provide valuable information that will make the system/processes better, if only one can learn from them. Like most things in life, if one can learn from a situation, then there is value. Focus on the purpose for the experiment and learn something.

For cases #2 and #3, because the two types of metrics had different results, it can provide a significant learning experience. In case #2, everything was done that the team set out to do and MOPs are green, but the expected impact was not accomplished. Does this mean that the tasks and objectives are not connected as previously thought? For case #3, even without completing all tasks, the intended effect occurred anyway. Was there something else that caused the result? Perhaps not every task needed to be completed to achieve the results.

In the gym example, case #1 is easy to describe. The subject went to the gym all the time, has lost the weight, and made the friends. Case #4 is equally easy. This subject signed up for the membership but has not set foot back in the gym, so they gained more weight and have lost the one friend they thought they had. In Case #2, the subject went to the gym and talked to people but is not losing any weight nor making any new friends. Why is that? Are they not working hard enough on the equipment? Are they not coming across as friendly? In case #3, the subject missed several workouts but is losing weight and made several friends. Are they doing something else that is not being measured in the MOPs, therefore causing the weight loss? In Case #5, this subject found that because people are hot and sweaty while working out, they are not associating with each other. This subject had more impact making friends by talking to people at the juice bar in the lobby.

11.2.5.2 Level of Measurement

Not all metrics are equal in the eyes of analysis and the way a question is asked will impact the data collected and in turn the type and depth of analysis that can be completed afterward. Data can be sorted in to four tiers: Nominal, Ordinal, Interval and Ratio. A different meaning can be extracted with each tier. Nominal data is typically categorical in nature, but in its basic form there is no sequence that each category should be placed in, such as blood type. Thus, minimal meaning can be extracted. For ordinal data there is a logical sequence, but relative size is meaningless, such as rating something poor/ fair/good/excellent, or age groups: infant, toddler, child, teenager…. The sequence is obvious, but how much better is good than fair? Interval data provides the answer to the range of difference between data points, such as temperature: 30 ºF, 60 ºF degrees and 90 ºF degrees are all exactly 30 ºF degrees different. So, there is true meaning in not just their sequence but also in the intervals between the points. However, 60 ºF is not twice as hot as 30 ºF. Interval data is stronger than ordinal, but it lacks a true zero and some mathematical operations cannot be performed. Ratio data on the other hand has a true zero. This is data such as age, weight, and height. To maximize the analysis that can be conducted afterward, it is always best to collect the highest tier data possible.

Using the example of a foot race, in a list of participants, nominal data would be the names of the schools represented by each runner. A histogram showing the number students represented by each school could be created but it would yield very little else. Ordinal data would show placement at the finish line, first/second/third. This data could provide ranking of the runners but would offer no insight for the span between the finish times for each of them. Was the result of the race a blow out or was it very close? Interval data is the amount of time between each runner, but if two runners crossed the finish line five seconds apart, is that a big gap? In a marathon where they run for hours, five seconds is almost nothing, but in a 100-meter sprint, five seconds is a lifetime. The ratio data for a race is the actual finish times. This data would enable the most complete analysis.

11.2.6 Data Collection Methods

There are many methods for collecting data within an experiment, ranging from completely manual to fully automated. The tools that are used to measure the metric and the way they are recorded should be established in advance and captured in the DCAP. What tool should be used depends on the nature of the data sets. For data sets that are very qualitative in nature a manual process might be preferred. With an expert in place to make the evaluation, a manual process can provide the details and insights needed to make an assessment. For a data set that is very quantitative in nature an automated process might be preferred. While some factors for the collection method might change by situation, it is normally a good idea for collections to be as discreet as possible and be done in a way that will ensure the integrity of the data.

11.2.6.0 Accuracy vs Precision

Accuracy refers to how close a measurement is to the true or accepted value. Precision refers to how close measurements of the same item are to each other. Precision is independent of accuracy. That means it is possible to be very precise but not very accurate, and it is also possible to be accurate without being precise.

A classic way of demonstrating the difference between precision and accuracy is with a shooting target. Think of the center of the target as the true value. The closer the shots land to the center, the more accurate they are. The tighter the cluster, the more consistent or precise. A tight grouping away from center is precise (but not accurate), a scattering that is uniform around the center is accurate but not precise. The best quality observations are both accurate and precise. If there is only one known, it is possible in the post processing to try to account for this but not always. For example, if all shots form a circle around one point, that could be accuracy but not precision. Averaging the values would provide a potential answer for the true value. On the other hand, a very tight grouping of shots says consistent or precise, but unless the target is known to be down and to the left, it will be difficult to fix the accuracy.

A key question that is often asked is how much accuracy and precision are needed, and the answer is, it depends on what is measured and the impact of being wrong. For example, the weight of a battleship is in tons, if one is off by a few pounds here or there the impact is minimal. If one is measuring the components for a satellite, the impact of being off by even a fraction of an ounce could mean being off balance and spinning out of control, with enormous impact.

11.3 Appendix C – Technology Readiness Level

As a new technology is developed, it starts as an idea in someone's head and eventually becomes a finished product. The defense acquisition community has developed a system to rate technology as it matures and provide a snapshot on its status, the Technology Readiness Levels (TRLs). The primary purpose of using TRLs is to help management in making decisions concerning the development and transitioning of technology. It should be viewed as one of several tools that are needed to manage the progress of research and development activity within an organization. Most exercises have a minimum TRL rating to be considered for inclusion, as shown in Figure 11-4. See the tables below[26] for descriptions of the TRLs and use to evaluate where the technology fits.

FIGURE 11-3 TRL LEVELS

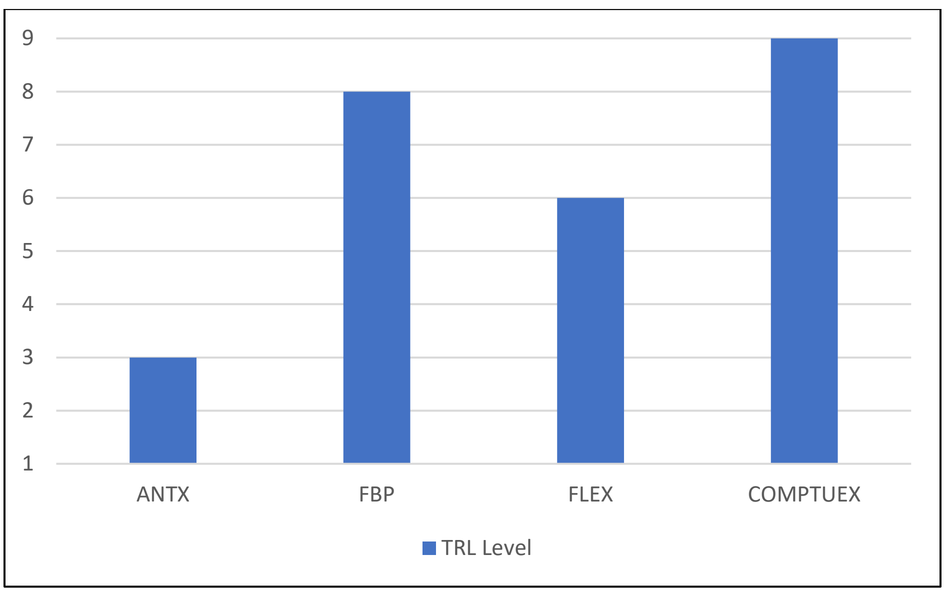

FIGURE 11-4 EVENT TRL LEVEL

[26] Technology Readiness Assessment Deskbook, 2009

Hardware TRL Definitions, Descriptions, and Supporting Information		
TRL Definition	**Description**	**Supporting Information**
1 *Basic principles observed and reported.*	Lowest level of technology readiness. Scientific research begins to be translated into applied research and development (R&D). Examples might include paper studies of a technology's basic properties.	Published research that identifies the principles that underlie this technology. References to who, where, when.
2 *Technology concept and/or application formulated.*	Invention begins. Once basic principles are observed, practical applications can be invented. Applications are speculative, and there may be no proof or detailed analysis to support the assumptions. Examples are limited to analytic studies.	Publications or other references that outline the application being considered and that provide analysis to support the concept.
3 *Analytical and experimental critical function and/or characteristic proof of concept.*	Active R&D is initiated. This includes analytical studies and laboratory studies to physically validate the analytical predictions of separate elements of the technology. Examples include components that are not yet integrated or representative.	Results of laboratory tests performed to measure parameters of interest and comparison to analytical predictions for critical subsystems. References to who, where, and when these tests and comparisons were performed.
4 *Component and/or breadboard validation in a laboratory environment.*	Basic technological components are integrated to establish that they will work together. This is relatively "low fidelity" compared with the eventual system. Examples include integration of "ad hoc" hardware in the laboratory.	System concepts that have been considered and results from testing laboratory-scale breadboard(s). References to who did this work and when. Provide an estimate of how breadboard hardware and test results differ from the expected system goals.
5 *Component and/or breadboard validation in a relevant environment.*	Fidelity of breadboard technology increases significantly. The basic technological components are integrated with reasonably realistic supporting elements so they can be tested in a simulated environment. Examples include "high-fidelity" laboratory integration of components.	Results from testing a laboratory breadboard system are integrated with other supporting elements in a simulated operational environment. How does the "relevant environment" differ from the expected operational environment? How do the test results compare with expectations? What problems, if any, were encountered? Was the breadboard system refined to more nearly match the expected system goals?
6 *System/subsystem model or prototype demonstration in a relevant environment.*	Representative model or prototype system, which is well beyond that of TRL 5, is tested in a relevant environment. Represents a major step up in a technology's demonstrated readiness. Examples include testing a prototype in a high-fidelity laboratory environment or in a simulated operational environment.	Results from laboratory testing of a prototype system that is near the desired configuration in terms of performance, weight, and volume. How did the test environment differ from the operational environment? Who performed the tests? How did the test compare with expectations? What problems, if any, were encountered? What are/were the plans, options, or actions to resolve problems before moving to the next level?
7 *System prototype demonstration in an operational environment.*	Prototype near or at planned operational system. Represents a major step up from TRL 6 by requiring demonstration of an actual system prototype in an operational environment (e.g., in an aircraft, in a vehicle, or in space).	Results from testing a prototype system in an operational environment. Who performed the tests? How did the test compare with expectations? What problems, if any, were encountered? What are/were the plans, options, or actions to resolve problems before moving to the next level?
8 *Actual system completed and qualified through test and demonstration.*	Technology has been proven to work in its final form and under expected conditions. In almost all cases, this TRL represents the end of true system development. Examples include developmental test and evaluation (DT&E) of the system in its intended weapon system to determine if it meets design specifications.	Results of testing the system in its final configuration under the expected range of environmental conditions in which it will be expected to operate. Assessment of whether it will meet its operational requirements. What problems, if any, were encountered? What are/were the plans, options, or actions to resolve problems before finalizing the design?
9 *Actual system proven through successful mission operations.*	Actual application of the technology in its final form and under mission conditions, such as those encountered in operational test and evaluation (OT&E). Examples include using the system under operational mission conditions.	OT&E reports.

TABLE 3 HARDWARE TRL LEVEL DESCRIPTIONS

Software TRL Definitions, Descriptions, and Supporting Information		
TRL Definition	**Description**	**Supporting Information**
1 *Basic principles observed and reported.*	Lowest level of software technology readiness. A new software domain is being investigated by the basic research community. This level extends to the development of basic use, basic properties of software architecture, mathematical formulations, and general algorithms.	Basic research activities, research articles, peer-reviewed white papers, point papers, early lab model of basic concept may be useful for substantiating the TRL.
2 *Technology concept and/or application formulated.*	Once basic principles are observed, practical applications can be invented. Applications are speculative, and there may be no proof or detailed analysis to support the assumptions. Examples are limited to analytic studies using synthetic data.	Applied research activities, analytic studies, small code units, and papers comparing competing technologies.
3 *Analytical and experimental critical function and/or characteristic proof of concept.*	Active R&D is initiated. The level at which scientific feasibility is demonstrated through analytical and laboratory studies. This level extends to the development of limited functionality environments to validate critical properties and analytical predictions using non-integrated software components and partially representative data.	Algorithms run on a surrogate processor in a laboratory environment, instrumented components operating in a laboratory environment, laboratory results showing validation of critical properties.
4 *Module and/or subsystem validation in a laboratory environment (i.e., software prototype development environment).*	Basic software components are integrated to establish that they will work together. They are relatively primitive with regard to efficiency and robustness compared with the eventual system. Architecture development initiated to include interoperability, reliability, maintainability, extensibility, scalability, and security issues. Emulation with current/legacy elements as appropriate. Prototypes developed to demonstrate different aspects of eventual system.	Advanced technology development, stand-alone prototype solving a synthetic full-scale problem, or standalone prototype processing fully representative data sets.
5 *Module and/or subsystem validation in a relevant environment.*	Level at which software technology is ready to start integration with existing systems. The prototype implementations conform to target environment/interfaces. Experiments with realistic problems. Simulated interfaces to existing systems. System software architecture established. Algorithms run on a processor(s) with characteristics expected in the operational environment.	System architecture diagram around technology element with critical performance requirements defined. Processor selection analysis, Simulation/Stimulation (Sim/Stim) Laboratory buildup plan. Software placed under configuration management. Commercial-of-the-shelf/government-off-the-shelf (COTS/GOTS) components in the system software architecture are identified.
6 *Module and/or subsystem validation in a relevant end-to-end environment.*	Level at which the engineering feasibility of a software technology is demonstrated. This level extends to laboratory prototype implementations on full-scale realistic problems in which the software technology is partially integrated with existing hardware/software systems.	Results from laboratory testing of a prototype package that is near the desired configuration in terms of performance, including physical, logical, data, and security interfaces. Comparisons between tested environment and operational environment analytically understood. Analysis and test measurements quantifying contribution to system-wide requirements such as throughput, scalability, and reliability. Analysis of human-computer (user environment) begun.
7 *System prototype demonstration in an operational high-fidelity environment.*	Level at which the program feasibility of a software technology is demonstrated. This level extends to operational environment prototype implementations, where critical technical risk functionality is available for demonstration and a test in which the software technology is well integrated with operational hardware/software systems.	Critical technological properties are measured against requirements in an operational environment.
8 *Actual system completed and mission qualified through test and demonstration in an operational environment.*	Level at which a software technology is fully integrated with operational hardware and software systems. Software development documentation is complete. All functionality tested in simulated and operational scenarios.	Published documentation and product technology refresh build schedule. Software resource reserve measured and tracked.
9 *Actual system proven through successful mission-proven operational capabilities.*	Level at which a software technology is readily repeatable and reusable. The software based on the technology is fully integrated with operational hardware/software systems. All software documentation verified. Successful operational experience. Sustaining software engineering support in place. Actual system.	Production configuration management reports. Technology integrated into a reuse “wizard.”

TABLE 4 SOFTWARE TRL LEVEL DESCRIPTIONS

11.4 Appendix D – Installation Processes

Each experiment requires some sort of engineering rigor and risk management process that should be followed. Several risk assessment processes exist, based on the experiment's risk to the ship and what must be done to mitigate it. For example, the fewer DoD components used in an experiment, the easier the process. Similar assessments and testing are applied to aircraft and submarines to understand and mitigate any risks and impacts presented by the experiment. Many of these tests and planning considerations are **common to all experiments or technical demonstrations whether they are on surface ships, shore facilities, aircraft, or submarines**. As stated above, not all of these will be required for every experiment; but the following is a list of considerations likely to be required for any fleet experiment or technical demonstration.

- The **Navy Risk Management Framework (RMF) for Cybersecurity** applies to all systems – without exception – that receive, process, store, display, or transmit DoD information, including systems participating in Navy experimentation or technical demonstrations with the goal of obtaining Interim Authorization to Test (IATT) prior to the install date for the event. A streamlined RMF process for experimentation has been developed to achieve IATT authorizations and fulfill Cybersecurity requirements using best practices from DoD partners and the Center for Internet Security (CIS) with the goal of improving RMF IATT processing times in support of experimentation requirements and timelines. The streamlined process may not be guaranteed in every circumstance, so the emphasis should be on the earliest possible start for RMF processing to avoid having the experiment stopped due to lack of IATT.[27]
- **Application Integration (AI) Assessment**: SBIR communities should be aware that early planning is crucial for acceptance into the lab testing environment; there are criteria and cyber accreditation requirements needed prior to lab environment entry. AI assessment is required for computing hardware or software integrated on any afloat network. The process is started through submission of the Afloat Service Request Form (SRF). It is then scheduled and executed in government labs with government personnel assisted by commercial submitters. PMW 160 holds sponsorship for the legacy and CANES networks but there are several trusted agents with the ability to accomplish this assessment. [28]
- **Mission Readiness Assessment (MRA)/Combat Systems Integrated Testing**: Assessment and testing provides evidence that systems, software applications, and hardware are functioning properly. It is important to note that if the experiment impacts the Integrated Combat System (ICS), the submitter (sponsor, PARM, or assignee) needs to enter it into the ICS Configuration Control Board (CCB) **before** formal submittal of the SCD through the Navy Data Environment (NDE).
- **Weapon System Explosive Safety Review Board (WSESRB)**: Initial installation testing, qualification testing, physical fit checks, status ground fire testing, systems integration lab (SIL), safety analysis, safe separation test certification, and Non-Nuclear Munitions Safety Board (NNMSB) may be gathered for review and concurrence through this board.
- **Ship Checks and Shore Site Visits**: These are performed in conjunction with the planning yard and the sponsor. Tasking and funding must be in place before they can begin. Ship Installation Drawings (SIDs), if needed, will be developed from information obtained through the Ship Check.[29]
- **Standard Frequency Action Format (SFAF)**: This may be required because some systems receive but do not transmit a signal. Electromagnetic Spectrum (EMS) support is often not required for these receive-only systems; however, these systems can be vulnerable to emissions from other devices. Standard frequency action format (SFAF) records are a way to identify the location of these devices for their protection as receive-only systems.[30]

---

[27] Appendix Q, Fleet Experimentation and Technology Demonstration, p. 23
[28] PMW 160 Fact Sheet
[29] Appendix Q, Fleet Experimentation and Technology Demonstration, SL720-AA-MAN-030, p. Q-23
[30] CJCSM 3320.01C, Joint Electromagnetic Spectrum Management Operations in the Electromagnetic Operational Environment, Enclosure C, p. C-6

A typical timeline can be seen in Figure 11-5. For more information and descriptions, consult the forthcoming Shipboard Installations Guidebook. FLEX and TECH DEMOs have a slightly different set of processes and requirements. Surface ships and unmanned surface vehicles (USVs) will follow the FLEX or TECH DEMO process described below.

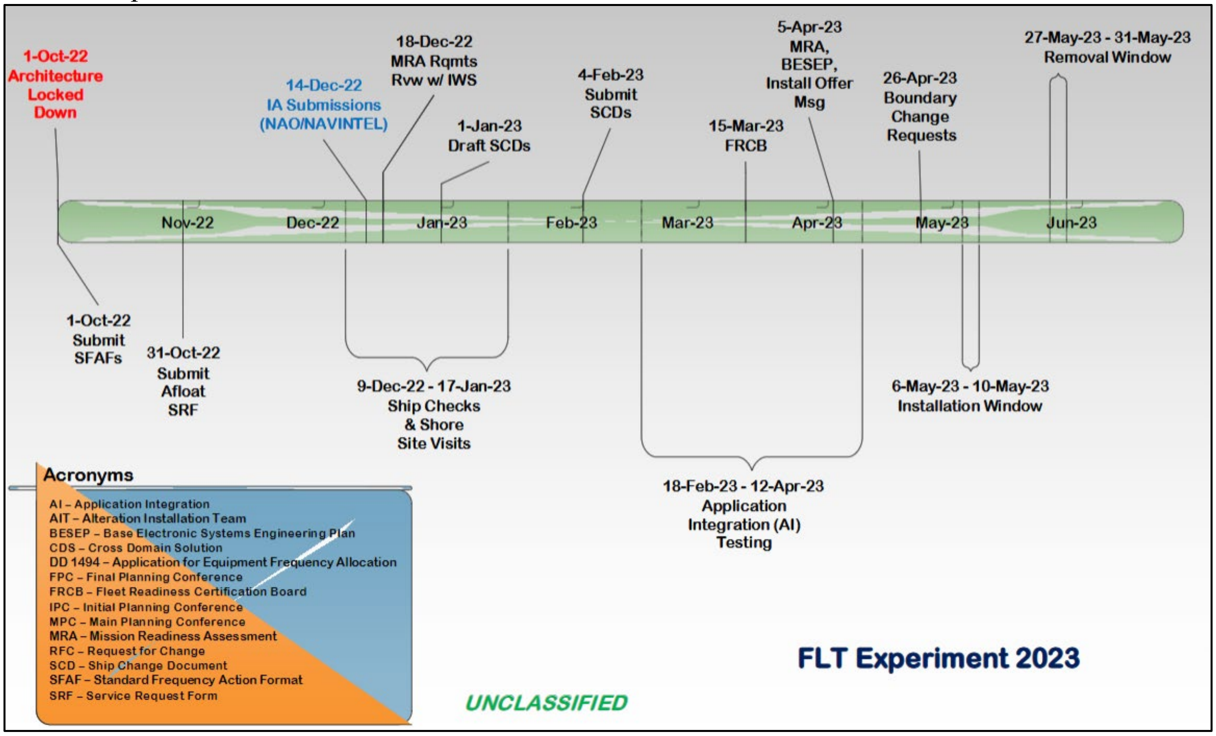

FIGURE 11-5 SAMPLE INSTALLATION TIMELINE

11.4.1.0 Surface Ships and AEGIS Ashore (FLEX or TECH DEMO streamlined NMP)

Both FLEX and TECH DEMOs will follow the streamlined NMP process for experimentation which either will fall into one or more of 12 installation types and four scenarios (or combinations of each). For FLEX, the experiment will be entered into the FLEX Information Management System (FIMS). A TECH DEMO will not go into FIMS; but will be entered into the Navy Data Environment (NDE). Each of the 12 installation types and scenarios will generate a set of requirements and deliverables that must be met to mitigate risk to the ship. These must be accomplished before execution of the experiment or demonstration on the ship. Low complexity experiments will generate a short list of requirements and high complexity experiments will generate a longer list of requirements.

For most installations that will occur on a ship for the purpose of experimentation or technical demonstration, a ship change document (SCD) will be required. Whether or not a full SCD is necessary will be determined by the type of installation and its level of complexity. The SCD will be filled out in NDE and will serve as a platform for the technical data package (TDP) and produce a tracking number for coordination and communication in relation to reviews and approvals. Almost all TECH DEMOs will need a SCD and most FLEXs will need one in addition to entry into FIMS. In the case of a less complex installation, a header-only SCD will be set up mainly to obtain a tracking number. The timeframe for approval will vary depending on the program (from an expedited 30 days to as many as 90) so it is crucial for the sponsor or a designated preparer to submit the SCD as early as possible. Open communication with key parties can be beneficial in a smooth approval process.

Fleet experiments or technical demonstrations fall into three categories: 1) High complexity requiring submittal of a full SCD through the NDE along with an accompanying TDP; 2) Moderate complexity requiring a "header-only" SCD in NDE for the purpose of obtaining a SCD number and setting up an accompanying TDP; and 3) Low complexity requiring only a Risk Assessment Request Message be sent for fleet commander approval. Some or all the following tests and assessments may be required and when complete, the documentation will then become a part of the TDP that accompanies the SCD.[31]

- The **Navy Risk Management Framework (RMF) for Cybersecurity**
- **Application Integration (AI) Assessment**
- **Mission Readiness Assessment (MRA)/Combat Systems (CS) Integrated Testing**
- **Total Ship Computing Environment (TSCE)/ Non-C4I Network Assessment/Test**: Required to discover whether systems, hardware, and software applications can operate with Non-C4I networks. It can be accomplished through desktop assessments, either shore-based or shipboard, and is coordinated through the PARM.[32]
- Installation plans, architecture drawings, outline, and installation (O&I) drawings: These plans and drawings are normally prepared by the PARMs.
- **Ship Checks**: If Ship Installation Drawings are required, these will be formally approved by the NAVSEA Planning Yard (PY) or authorized third-party and will be developed from the results of the Ship Check.[33]
- **Topside Assessment**: This assessment examines any equipment that may be installed topside that could impact the electromagnetic performance of the ship's existing systems. It needs to occur very early, possibly **one year before the experimentation event**. Guidance may be obtained from the Integrated Topside Design-Surface Ships Technical Warrant Holder (TWH).[34]
    - EMI 461 Assess/Test: SECNAVINST 5000.2E – This is an E3 requirement for electrical and electromagnetic systems or equipment that will be installed on a ship to ensure they are compliant at the box and subsystem level for electromagnetic interference compatibility for both the equipment/system and for other systems on the ship.[35]
    - EMI 464 Assess/Test: Electromagnetic Effects (E3): MIL-STD-464C- E3 Requirements for Systems that relates to the risks imposed on ships and platforms from electromagnetic and radio frequency interference by antennas or other emitters. Approved locations for this type of equipment on the ship will be determined by this assessment.[36]
    - HERO/HERF/HERP: These tests find possible hazards of electromagnetic radiation: Hazards to Ordnance (HERO); Hazards to Fuel (HERF); and Hazards to Personnel (HERP).[37]
- **Frequency Spec Certification**: This is required for a software installation where an experiment will have an impact on an existing frequency spectrum (portable and non-portable transmitting antennas).[38]
- **Lithium Battery Certification**: A certification letter must be obtained from the Lithium Battery Safety Program for systems or devices that contain a lithium battery.
- **Laser Safety Hazards Control** (Laser Safety Review Board (LSRB)): Review is required for laser systems used in combat, combat training, and those capable of exceeding Class 3R, 3B and 4 levels.

[31] Appendix Q, Fleet Experimentation and Technology Demonstration, SL720-AA-MAN-030, p. Q-12
[32] Appendix Q, Fleet Experimentation and Technology Demonstration, p. 25
[33] Appendix Q, Fleet Experimentation and Technology Demonstration, p. 23
[34] Appendix Q, Fleet Experimentation and Technology Demonstration, p. Q-55
[35] Department of Defense Guidebook for Electromagnetic Spectrum Survivability, p. 30
[36] MIL-STD-464C
[37] Appendix Q, Fleet Experimentation and Technology Demonstration, p. Q-57
[38] OPNAVINST-2400.20, p. 7

This includes those used in optical fiber communication systems and will ensure compliance with regulations, the recording of laser hazard data, and recommended eye protection safety.[39]
- FLEX or TECH DEMO Risk Assessment Request Message (RA REQ MSG): The naval message will be used to obtain approval from the FLTCDR for the installation (rather than a SCD). The message must be sent using the Navy Message Format and is normally sent by the sponsor or assignee.[40]
- RMMCO Check In/Out: When an experiment is complex and requires a SID and/or is industrial in nature, installation teams will be required to observe the Check in/Check process out through the Regional Maintenance and Modernization Coordination Office (RMMCO).
- **FLEX or TECH DEMO Removal Message** (REM MSG): This message is used to report removal of a FLEX installation and is normally sent by the sponsor or assignee.[41]

The following tests are rarely needed, but are listed here for situational awareness:
- **Shock Assess Test**: Either shock testing or an approved Shock Deficiency Correction Plan (SDCP) may be required for hardware or firmware that will be installed on the ship or the modification of any internal equipment. Bottom Line Up Front (BLUF) is used for testing. Guidance may be obtained from the Shock Technical Warrant Holder (TWH).[42]
- **Vibration Assessment/Test**: Required for equipment wherein failure modes could cause a hazard to the ship, equipment, or personnel.[43]
- **Fire, Smoke and Toxicity Tests**: May be required for non-metallic technology insertion.

11.4.1.1 Aircraft (Airworthiness Certification for Interim Flight Clearance)

In addition to those requirements listed above for ALL experiments, the following information can help discover additional requirements and processes for obtaining airworthiness certification and interim flight clearances. AIR-4.0P, the CYBERSAFE Directorate, owns the process. Experimental airworthiness certificates may be issued for research and development activities and limitations may be imposed on operations and maintenance of the aircraft.[44] There are two types of Interim Flight Clearances:
- Interim Flight Clearance (IFC) through the automated web request system and using the Naval Aviation Technical Information Product (NATIP).
- The Naval Air Training and Operating Procedures Standardization (NATOPS) permanent flight clearance (PFC) through the Airworthiness Issue Resolution System (AIRS). These are submitted by NAVAIR personnel.

The following steps must be observed:
- Flight clearance planning meeting to determine the scope of the IFC.
- Data must be provided to Technical Area Experts (TAEs) through the planning meeting.
- AIR-4.0P will come to a decision.
- TYCOM, ACC, or Program Office will concur with the request if it is for an IFC; or a TYCOM concurrence may have been pre-coordinated (with NAVAIR ARC) for Test clearance requests submitted by the AIR-4.0P Test Flight Clearance.[45]

---
[39] NAVSEA S9070-AA-MME-010/SSN/SSBN, Revision 3, CAN-5, p. 8
[40] Appendix Q, Fleet Experimentation and Technology Demonstration, pp. Q-70 - 86
[41] Appendix Q, Fleet Experimentation and Technology Demonstration, pp. Q-87 - 88
[42] Appendix Q, Fleet Experimentation and Technology Demonstration, p. Q-54
[43] FY21 NAVSEA Standard Item No. 009-104
[44] NAVAIR MANUAL M-13034.1, pp. 1-3 and 3-7
[45] NAVAIR MANUAL M-13034.1, pp. 3-12

11.4.1.2 Unmanned Aircraft Systems (Airworthiness Certification for Interim Flight Clearance)

There are three categories of airworthiness:

- The Unmanned Aircraft System (UAS) is already designed and qualified to the DON standards or DON equivalent.
- The UAS level of airworthiness needs to correlate to the Probability of Loss of Aircraft (PLOA) rate of no more than 1 loss per 10,000 flight hours, less than or equal to 12,500 lbs.
- The UAS is assessed as a set of airworthiness criteria and controlled by stringent operating limitations in restricted areas, warning areas, combat zones, U.S. National airspace, international airspace, and shipboard launch and recovery. A UAS in this situation will require greater engineering review, especially when the flight occurs from/to a ship. When a flight is planned within U.S. Airspace, a Certificate of Waiver and Authorization is required for flights outside of a restricted area or warning area.[46]

Additional Airworthiness Certification Criteria (MIL-HDBK-516C) may be required. Not all criteria will apply to every air system and platform; unique criteria may need to be added.[47]

- Design Criteria may be needed to verify safety for usage and permissible flight envelope.
- Failure Conditions may need to be verified and addressed.
- Tools and Databases may require verification that methods are applied appropriately, and information is running through the correct databases.
- WSESRB concurrence may be required through initial installation testing, qualification testing, physical fit checks, status ground fire testing, SIL, safety analysis, safe separation test certification, NNMSB[48]
- CNS/ATM certification is required for Research, Development, Test and Evaluation (RDT&E) activities that occur on rotary wing avionics mission systems.[49] [50]

11.4.1.3 Submarine (Temporary Submarine Alterations (TEMPALT))

The Supervisor of Shipbuilding (SUPSHIP) will need to be notified of a submarine TEMPALT. SUPSHIP assigns the TEMPALT number and contact should be maintained throughout the process. SUPSHIP will require payment for review of the TEMPALT, so it is important to have funding in place first. In addition to those requirements listed above for ALL experiments, other possible requirements for submarines are as follows: [51]

- Accreditation Package Development/Risk Management Framework (RMF): In addition to the streamlined process that is in place entitled, “NAO – Streamlined Process for Experimentation 2020” described above, submarines will go through a TEMPALT (TA) Cybersecurity Evaluation.[52]
- EMI Surveys: In addition to the EMI considerations described above, EMI surveys for submarines cannot occur in a manner that would impede forward sonar and communications systems access or cause a power-down of systems, unless notification has been provided ahead of time to the EMC technician. To obtain an accurate assessment, forward electronics must be energized like the possible at-sea lineup. EMI surveys must be conducted by NAVSEA or NAVSEA designates.[53]

---

[46] NAVAIR MANUAL M-13034.1, pp. 6-3 - 6-5
[47] MIL-HDBK-516C, p. 2
[48] MIL-HDBK-516C, p. 490
[49] MIL-HDBK-516C, pp. 49, 51, 114, 395-396, and 489-490
[50] NAVAIR MANUAL M-13034.1, pp. 2-2 and 3-7
[51] SUPSHIP Operations Manual (SOM), S0300-B2-MAN-010 Rev 2, Change #21, Chapter 10
[52] PMS392 TempALT Submission Checklist, p. 1
[53] Joint Fleet Maintenance Manual, Vol. VI, 4.3.2.2

- Other TEMPALT items to consider are found on the *PMS 392 TEMPALT Submission Checklist* and the *Technical Requirements Manual for Temporary Submarine Alterations.* Both sources are not generally viewable by the public, but the sponsor will likely be able to access them.

The processes for these areas will be tailored toward the unit or domain in which the system will be installed, most have similar guidelines and sub-processes. As discussed above, some processes and assessments are common to any type of experimentation or demonstration. These are listed to provide some help in forecasting possible requirements and the timelines involved to move more efficiently through the processes.