Enabling Multi-Domain Operations Through Wargames, Simulation, and Live Exercises

Per-Idar Evensen, Helene Rødal Holhjem, Daniel Myklatun Tveit, Karolina Di Remigio Eikås
Norwegian Defence Research Establishment (FFI)
Kjeller, Norway
per-idar.evensen@ffi.no, helene-rodal.holhjem@ffi.no, daniel-myklatun.tveit@ffi.no, karolina-di-remigio.eikas@ffi.no

ABSTRACT

The emerging operational concept of multi-domain operations (MDO), which will require a high degree of integration and synchronization of capabilities and activities across all the operational domains (land, maritime, air, space, and cyberspace) and a high operational tempo, needs to be further developed, detailed, tested, and validated through experimentation and analysis. Experimentation and analysis provide the ability to iteratively explore, test, refine, and validate concepts. Furthermore, modeling and simulation (M&S) is essential for being able to experiment with, and analyze the effectiveness of, different concepts.

Effective MDO will rely upon the ability to effectively adapt new technologies and novel thinking into increased combat effectiveness to deal with a rapidly changing operating environment. However, it is often challenging to implement a coherent and holistic end-to-end approach for concept development from ideas to fully functional and fielded concepts. This is especially true when it comes to concept development involving stakeholders and forces from all, or several of, the five operational domains.

At the Norwegian Defence Research Establishment (FFI), we have had success with a methodology based on a structured, data-driven process of wargaming, simulation, and live exercises that follows ideas from conception to fielded concepts, thereby enabling rapid iteration of concept development and driving the Lessons Learned Process (LLP) from observation to implementation and verification. We have used the methodology at the joint level and are now adapting it for concept development for MDO. Furthermore, we envisage moving toward a data-centric approach that facilitates structured data collection and analysis in each step of the process so the data can be more effectively managed, accessed, and utilized.

In this paper we describe our methodology for concept development through iterations of wargames, simulation, and live exercises, along with examples of successful use cases, and discuss challenges encountered and possible future improvements of the methodology.

ABOUT THE AUTHORS

Per-Idar Evensen is a Principal Scientist at the Norwegian Defence Research Establishment (FFI), where he has worked since 2006. His current research work focuses on modeling and simulation (M&S) of multi-domain operations at different levels for experimentation and analysis purposes, and he has published more than a dozen peer-reviewed papers in M&S-related conference proceedings and journals. Per-Idar has a Master of Science degree in Computer Science from the University of Oslo in 2004.

Helene Rødal Holhjem is a Principal Scientist at the Norwegian Defence Research Establishment (FFI), where she has worked since 2007. Her current research work focuses on using modeling and simulation (M&S) and simulation-based experiments to support development of operational concepts. She has a Master of Science degree in Applied Physics from the Norwegian University of Science and Technology in 2007.

Daniel Myklatun Tveit, Ph.D., is a Senior Scientist at the Norwegian Defence Research Establishment (FFI). His research is focused on utilizing modeling and simulation to support experimentation for doctrine development and testing, and to facilitate and analyze live exercises and operations. Daniel was awarded a PhD in control theory and systems biology from the University of Stavanger in 2020. Daniel has been heavily engaged with interdisciplinary research, in areas ranging from control engineering and signal processing to systems biology and cancer metabolism.

Karolina Di Remigio Eikås, Ph.D., is a Scientist at the Norwegian Defence Research Establishment (FFI), where she has worked since 2024. Her research is focusing on modeling and simulation of multi-domain operations. Karolina has a PhD in theoretical chemistry from UiT – The Arctic University of Norway in 2022.

Enabling Multi-Domain Operations Through Wargames, Simulation, and Live Exercises

Per-Idar Evensen, Helene Rødal Holhjem, Daniel Myklatun Tveit, Karolina Di Remigio Eikås
Norwegian Defence Research Establishment (FFI)
Kjeller, Norway
per-idar.evensen@ffi.no, helene-rodal.holhjem@ffi.no, daniel-myklatun.tveit@ffi.no, karolina-di-remigio.eikas@ffi.no

INTRODUCTION

The emerging operational concept of *multi-domain operations* (MDO), which will require a high degree of integration and synchronization of capabilities and activities across all the operational domains (land, maritime, air, space, and cyberspace) and a high operational tempo, needs to be further developed, detailed, tested, and validated through experimentation and analysis. Experimentation and analysis provide the ability to iteratively explore, test, refine, and validate concepts. Furthermore, modeling and simulation (M&S) is essential for being able to experiment with, and analyze the effectiveness of, different concepts.

Effective MDO will rely upon the ability to effectively adapt new technologies and novel thinking into increased combat effectiveness to deal with a rapidly changing operating environment. However, it is often challenging to implement a coherent and holistic end-to-end approach for concept development from ideas to fully functional and fielded concepts. This is especially true when it comes to concept development involving stakeholders and forces from all, or several of, the five operational domains.

At the Norwegian Defence Research Establishment (FFI), we have had success with a methodology based on a structured, data-driven process of *wargaming*, *simulation*, and *live exercises* that follows ideas from conception to fielded concepts, thereby enabling rapid iteration of concept development and driving the Lessons Learned Process (LLP) from observation to implementation and verification. We have used the methodology at the joint level and are now adapting it for concept development for MDO. Furthermore, we envisage moving toward a data-centric approach that facilitates structured data collection and analysis in each step of the process so the data can be more effectively managed, accessed, and utilized.

In this paper we first describe the background for this work. Then, we describe in detail our methodology for concept development through iterations of wargames, simulation, and live exercises. Moreover, we outline a detailed example of a successful use case with a concept for a coordinated missile attack from forces in the air and maritime domains. Finally, we discuss challenges encountered and possible future improvements of the methodology.

BACKGROUND

Various forms of wargames and computer-based combat simulations have been carried out for concept development, experimentation, and analysis at FFI, for decades. Seminar-type, discussion-based wargames have often been used for exchanging, challenging, and developing new ideas, especially in the early phases of a concept development process. Simulation-based experiments have often been used to assess and compare the performance and combat effectiveness of different, more mature, concepts (Martinussen et al., 2008; Hoff et al., 2012; 2013).

In the last few years, we have employed a structured, data-driven process of wargaming, simulation, and live exercises that follows ideas from conception to fielded concepts. We have used this methodology for concept development at the joint level. The North Atlantic Treaty Organization (NATO) has chosen MDO as its future warfighting concept. Consequently, there is a shift in Norway toward MDO. We are therefore adapting the methodology for concept development for MDO.

MDO is an overarching concept for future operations where the underlying idea is seamless integration of capabilities and activities across all operational domains to present the enemy with multiple simultaneous dilemmas and achieve overwhelming local superiority in time and space on the battlefield. NATO's definition of MDO is: "The orchestration of military activities, across all operational domains and environments, synchronized with non-military activities, to enable the Alliance to create converging effects at the speed of relevance" (NSO, n.d.). The concept of MDO can be seen as a natural evolution of joint operations (NSO, 2022). MDO places increased emphasis on the space and cyberspace domains. MDO is also more focused on seamless integration and synchronization of capabilities from all domains at all levels of warfare and especially at the tactical/engagement level, increased operational tempo, and synchronization with non-military activities.

The Norwegian Long-term Defence Plan 2025–2036 (Norwegian MoD, 2024) states that "The Norwegian Armed Forces shall support the alliance's ability to [conduct] multi-domain operations". Furthermore, it highlights that the two most important research and development (R&D) priorities for the Norwegian defense sector are that "Norway will *lead and participate in multi-domain operations* and *be at the forefront among allies in situational awareness in the North.*"

FFI's strategy for 2044 (FFI, 2024) has MDO as one of five (multi-domain operations, situational awareness, data-driven defense, total defense, and climate and environment) important areas of research, development, and innovation (R&D&I). Regarding MDO, the strategy states that "FFI aims to conduct R&D&I activities that enable the [Norwegian] Armed Forces to operate effectively in all domains, both independently and alongside our allies. This will require superior situational awareness and data-driven defences" (FFI, 2024). Priority areas for R&D&I within MDO are (FFI, 2024):

- Concepts and technological and structural development
- Weapons and weapons systems
- Effective, secure, and prompt operational coordination within the Norwegian Armed Forces, between civil society and the military, and with Norway's allies

In January 2025, FFI's strategy materialized in the initiation of a scientific prioritization of MDO through the project "Multi-Domain Operations – Concepts, Technology- and Structure Development". The scientific prioritization of MDO will contribute to developing and accelerating the Norwegian Armed Forces' ability to implement MDO, based on NATO's MDO concept, by demonstrating how new concepts, new structural elements, and new technology can make the Norwegian Armed Forces operate more effectively. The Norwegian Armed Forces' contribution to conflict prevention in Norway's immediate areas is not created by individual capabilities, but by the Armed Forces' combined resources. The Norwegian Armed Forces must therefore have the ability to coordinate and concentrate efforts with multiple capabilities simultaneously. The Norwegian Armed Forces must also support NATO's ability to conduct operations in the land, maritime, air, space, and cyberspace domains. Topics that will be covered by the scientific prioritization of MDO include:

- The importance of space and cyberspace operations in MDO
- The importance of communication in MDO
- Planning and command and control (C2) of MDO in Norway's surrounding areas
- New technology – platforms, methods, and processes
- Quality goals and learning effect in MDO

METHODOLOGY FOR CONCEPT DEVELOPMENT

Experimentation and analysis are key enablers for concept development and provide the ability to iteratively explore, test, refine, and validate concepts (NATO ACT, 2021). The experiments we conduct can be categorized as *discovery experiments* or *hypothesis-testing experiments*. *Discovery experiments* involve introducing novel systems, concepts, organizational structures, technologies, or other elements to a setting where their use can be observed and catalogued (Alberts & Hayes, 2002). *Hypothesis-testing experiments* are used to advance knowledge by seeking to falsify specific hypotheses or discover their limiting conditions (Alberts & Hayes, 2002). Wargaming and M&S are essential for experimentation with, and further development and detailing of, the MDO concept.

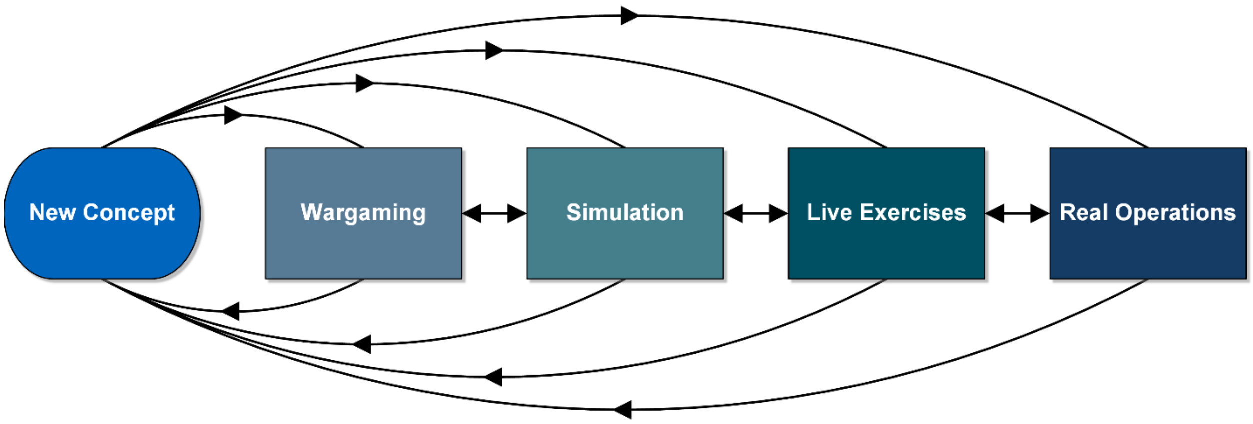

Figure 1. Concept Development Methodology.

At FFI we have successfully employed a concept development methodology based on an iterative, structured, data-driven process of wargaming, simulation, and live exercises that follows ideas from conception to fielded concepts. This process is illustrated in Figure 1 and is similar to the concept-to-capability experimentation campaign described in Numrich et al. (2017) and Numrich & Woods (2024).

Wargaming, combat simulation, and live military exercises are all forms of simulated representation of warfare. In this paper, however, we specifically use the term *wargaming* to refer to traditional wargaming that is carried out manually, with manual adjudication, and we specifically use the term *simulation* to refer to computer-based simulation, with and without humans in the loop.

The concept development process is usually initiated by an identified capability gap that needs to be closed or a new idea of how some type of task or mission can be done more effectively. New concepts can include the use of *new technologies*, *new ways of operating,* or *new organizational structures*. Concepts may also combine these elements. For concept development within MDO, it is important to ask: *Can a specific type of task or mission be performed more effectively by including capabilities and effects from additional domains?* While concept development for MDO can occur at all levels of warfare, our primary focus is concept development at the operational and tactical levels, where integration and synchronization of capabilities, activities, and effects across domains must be planned, executed, and evaluated. Insights from our experiments may also inform strategic-level decisions regarding force structure, capability development, and doctrinal evolution. Examples of concept development needed for MDO include concepts for more effective command and control (C2); resilient communication; more flexible organizational arrangements and command relationships; use of new technologies, for example artificial intelligence (AI) for decision support and autonomous vehicles; handling increased complexity; more effective force employment; new tactics, techniques, and procedures (TTP); coordination and synchronization of activities; synchronization of effects from multiple domains; new combat structures; and new force structure elements. Our whole concept development process is conducted in close collaboration with relevant stakeholders and actors in the Norwegian Armed Forces.

Wargaming, simulation, and live exercises are the three most important tools we have in our toolbox for concept development for MDO. Furthermore, by following a concept development process that starts with wargames, continues with simulations, and then moves on to live exercise, the concept will gradually be refined, matured, concretized, and detailed, until it reaches the final stage as a fully fielded concept. The concept development process is iterative, and a concept may go several rounds between the phases before it reaches the final stage. Simulations can, for example, show that a concept has a fundamental weakness not revealed in the wargaming phase. The concept must then be fundamentally redesigned, and a new round of wargaming will usually be necessary. Furthermore, a fully fielded concept may be further developed with new improvements, and this will usually trigger a new iteration of the whole concept development process.

The concept development process is data-driven, and systematic data collection and analysis are important parts of each phase in the process. In general, the results and analysis from one experiment in the process will determine what will be the next step.

The three distinct phases in our concept development methodology are described in more detail in the subsections below.

Wargaming

Wargames for research and analysis are usually referred to as *analytical* or *analytic wargames*. The general purpose of analytical wargames is to gain insight into complex issues related to warfare, for example to identify opportunities, challenges, or unforeseen consequences. Analytical wargames often aim to explore emerging concepts, ideas, or capabilities and are essentially a form of discovery experiment. Wargame expert Peter Perla argues that the "active and central involvement of human beings is the characteristic that distinguishes wargames from other types of models and simulations" (Perla, 1990). Moreover, "a value unique to all [wargames] is the occurrence of previously unknown issues, insights, or decisions that arise during the conduct of a game" (Burns, 2015).

Our wargaming phase will typically start with *seminar-type, discussion-based analytical wargames*. This type of wargame has less structured gameplay but offers the highest flexibility and is useful for collaborative discussion, exploration, and qualitative analysis of the new concept. If necessary, the seminar wargames may be followed by more detailed wargames with physical models on a paper map, but usually with minimal reliance on detailed rules and procedures. This type of wargame is often referred to as *system wargames* or *Free Kriegsspiel*.

The collected data and results from an analytical wargame are only as good as its collective body of players (Appleget et al., 2020). The players or participants in our wargames are military personnel (usually officers) and other military subject-matter experts (SMEs). They are usually assigned different roles. It is important that all relevant stakeholders and actors related to the concept are represented. In a wargame for concept development for MDO, it is also important that all relevant domains are represented. The vignettes or scenarios we use in our wargames are usually derived from overarching national defense scenarios.

In all types of analytical wargaming for concept development, it is important to have a competent opposing, or Red, side that can challenge the new concept. The players on the opposing side should also have good knowledge of the doctrine of expected opponents (Evensen et al., 2019; 2021; 2022c).

Defining a *clear objective* for each wargaming experiment is essential. The objective states the purpose of the wargame, i.e., what we want to achieve from the wargame. This could for example be to explore how military personnel employ the new concept, to gain insight into the strengths and weaknesses of the new concept, or to answer specific research questions for refinement, concretization, and detailing of the concept. A clear purpose for the wargame must be specified early in the preparation phase and will be the basis for the design of the wargame.

We have had success with following the methodology described in Appleget et al. (2020) for creating, executing, and analyzing analytical wargames. Designing and preparing a wargame can take from several weeks to a couple of months, and the wargaming phase usually lasts between six and twelve months. One crucial part of designing and developing an analytical wargame is to create a *data collection and management plan*. The data collection and management plan must decompose the objective of the wargame, and related issues or research questions, into a list of *essential questions* that need to be answered to achieve the objective of the wargame.

The analysis data from our wargames are collected manually and include notes taken by a data collection team during the wargames (often using structured forms with the essential questions) and notes from the after-action review (AAR) sessions, in addition to answers from questionnaires for the players and in-depth interviews with the players after the wargames. The results and analysis after a wargame often reveal issues that require further investigation.

Wargames cannot be used to predict the outcome of a battle or war, but they can produce plausible outcomes. The term "indication" has been suggested to describe any insights drawn from the outcome of a wargame (Rubel, 2006; Caffrey Jr., 2019). The factors that can affect the validity and credibility of wargames are discussed in (Evensen et al., 2021).

When the concept has been sufficiently refined, concretized, and detailed to be represented in a simulation system, we move on to the simulation phase.

Simulation

Simulation experiments are well suited for testing and comparing the relative performance and effectiveness of different military force structure elements that can vary with respect to the composition of materials and equipment, including new technologies; tactical organization; and operational concept (Evensen et al., 2022a; 2022b). There are mainly two approaches for conducting combat simulations: using *human-in-the-loop (HITL) simulations* (with varying degrees of human interaction) or using fully automated *closed-loop simulations* (without any human interaction). *Closed-loop simulations* can be run faster than real-time and thus repeated many times to get a statistical distribution of the outcomes, but they give a less realistic representation of the human aspects of combat. *HITL simulations* must be run in (or close to) real-time and can therefore only be repeated a few times, but they give a more realistic representation of the human aspects of combat. HITL simulations are mainly associated with virtual simulations in the live, virtual, and constructive (LVC) taxonomy, but constructive simulations may also require a certain degree of human interaction, for example, to control semi-automated forces (SAF). Generally, we only use virtual simulations in experiments where human system operators are essential, for example to experiment with technologies or concepts that directly affect human performance or how humans operate at the technical level (Evensen et al., 2024).

Human-in-the-Loop Simulations

Our simulation phase will typically start with HITL simulation experiments. The purpose of the HITL simulation experiments is to test and explore the new concept in a more realistic synthetic environment. These experiments are conducted using distributed interactive simulations where military officers and other military SMEs participate as role-players and decision makers and execute their mission according to prepared plans and defined procedures. We typically use constructive simulations, with SAF controlled by the human role-players/operators, occasionally together with some virtual entities, like combat aircraft and other weapon platforms (Evensen et al., 2024). The virtual entities will then typically be controlled by system operators in externally connected simulators.

In contrast to the manual wargaming phase, HITL-simulations immerse the players in a synthetic environment where combat elements and capabilities are subject to modeled physical constraints, including platform performance, sensor effects, and terrain influence. Actions must be executed in real time, forcing players to make decisions under pressure, based on their local situational awareness, and while facing operational friction. This exposes challenges such as delays, misunderstandings, and coordination failures that are often absent in manual wargames. Unlike manual wargames, where SMEs may “handwave” outcomes, HITL simulations enforce physical constraints and doctrinal limitations consistently, helping to identify feasibility issues. They also support automated, high-fidelity data logging, enabling repeatable analysis, quantitative metrics, and rich AAR sessions. As such, HITL simulations serve as a critical transitional phase between manual wargaming and closed-loop simulation, where the concept matures from an abstract idea to a concrete, testable construct. HITL simulations allow the concept to be explored and tested in dynamic, time-sensitive, and human operator-driven vignettes, enhancing the rigor and credibility of the overall concept development process.

In our HITL simulation experiments, we are typically able to conduct one to two simulation runs per day, and an HITL experiment series typically lasts from one to several days. Data from the synthetic environment, and communications between the players, are logged in a database for each run. After each run the participants meet for an AAR session where the chosen course of action (COA) is presented and the players convey their actions, the reasons behind their actions, and other experiences during the execution. Any technical issues during execution of the HITL simulation are also important to communicate during these sessions, as such issues can influence the experiences and results from the experiment. Data logging during the HITL simulation experiment enables the collection of quantitative results. Combined with qualitative data in the form of feedback from military SMEs participating in the experiment, this provides a deeper insight into the concept’s strengths and weaknesses. Our HITL simulation experiments can also be classified as discovery experiments.

After the manual wargaming phase and the HITL simulations, we are typically left with a set of hypotheses about the new concept, or about alternative concepts, that require further investigation. These may include, for example, hypotheses regarding which of two or more platform constellations or operational concepts that offer the highest relative combat effectiveness. A hypothesis that should always be tested is whether the new concept is significantly more effective than the current solution (i.e., the null alternative). Investigating such research questions requires statistically significant results, and closed-loop simulations, due to their ability to be executed many times under controlled and repeatable conditions, are particularly well suited for this type of experimentation.

Closed-loop Simulations

HITL simulations can typically only be run a limited number of times, making it difficult to determine whether the observed outcomes are representative or statistical outliers. Closed-loop simulations, on the other hand, can be executed repeatedly to capture stochastic variance caused by probabilistic elements, allowing for robust statistical analysis of performance, effectiveness, and sensitivity to input parameters. Moreover, these simulations can be executed as fast as the available hardware and software allows. As a result, closed-loop simulations can collect statistically significant results with far fewer resources than HITL simulations. In closed-loop simulations human behavior and communication must be represented by simulation models. This can typically be achieved by translating successful actions and decisions observed during the wargaming phase and HITL experiments into rules, decision logic, or behavior trees, which are then implemented within the simulation tool. This approach helps ensure that the automated agents in closed-loop simulations reflect realistic tactical behavior based on earlier human decision-making, while enabling repeatable and scalable experimentation.

Closed-loop simulations can incorporate stochastic models and detailed representations of the physical properties of force elements and the synthetic environments in which they operate. By systematically varying different variables of the closed-loop experiment, the results from a set of simulation iterations can contribute to increased insight into the area of validity of the chosen concepts and highlight the cases where the concept needs to be adapted further to increase combat effectiveness. The results from closed-loop simulations may also generate new research questions which can be explored further by either HITL simulation experiments or a new wargaming phase. Our closed-loop simulations can be categorized as hypothesis-testing experiments.

Modeling complex operations like MDO using closed-loop simulations requires numerous simplifying assumptions, and the validity of such simulations can often be questioned (TTCP, 2006). Our methodology, which first uses wargaming and HITL simulations, gives a good starting point for credible closed-loop simulations since the technical aspects regarding weapons, sensors, and communications, and force employment aspects such as doctrine, tactics, and procedures have already been discussed thoroughly and to some extent been tested. Another important aspect of validating closed-loop simulation for warfighting experimentation is the distinction between statistical significance and relevance to military operational impact (Willis et al., 2023). Closed-loop simulations are ideal for detecting changes and isolating the cause (Kass, 2006), but it is important to ensure that the change(s) actually have operational impact, and that it is not only a statistically significant change.

In closed-loop simulations it is important to log all events, actions, entity states, and interactions between entities as a function of simulation time. Batch runs can, however, produce large volumes of data, which can be challenging to manage. For complex combat simulations, like MDO simulations, it can also be more challenging to ensure traceability between cause and effect across multiple iterations.

After the simulation phase, the natural next step is to validate elements of the new concept in one or more live exercises.

Live Exercises

Simulation is an important tool for military experimentation in cases where the live execution would be too difficult, too expensive, or too dangerous. However, simulations necessarily include simplifications and assumptions of the real operational environment. When developing new concepts, it is therefore important to test elements of the new concept in live exercises. Live exercises provide a higher level of realism, which can provide new insights into the concept development process.

In our concept development methodology, live exercises play a crucial role in providing practical, real-world experiences with the new concept. It is in this step that concepts are tested and validated within a complex, realistic environment, with full interactions, and under time pressure (i.e., “Train as you fight, fight as you train”). One important difference from the other steps in our concept development methodology is that live military exercises are not planned and facilitated by FFI. This is the responsibility of the Norwegian Armed Forces. Therefore, in order to ensure that testing and validation is included in the exercise program, our methodology requires close cooperation with exercise planners in the Norwegian Armed Forces. FFI engages with exercise planners to identify appropriate exercises for testing and validation and helps formulate training and exercise objectives, along with analysis requirements and objectives. Exercise planners are then able to include activities for testing and validation in the exercise program, while FFI develops analysis and data collection plans for the exercise.

During the exercise, personnel from FFI will attend either as observers or as a part of an exercise evaluation team. This way, we can ensure that critical data for testing and validation are logged and that we can respond to any changes during the exercise to ensure that all analysis objectives are met. System data is often logged locally on any given military system and might require manual setup for data collection, while other data types are shared on networks and can be collected automatically. Developing a detailed data collection plan and communicating the specific data requirements, as well as means of data collection, for the live exercise is therefore of great importance.

After the exercise, scientists from FFI organize the data collected to identify which analysis objectives can be answered. In the case of incomplete data, an attempt is made to acquire additional data post-exercise. The data is then analyzed, from which evidence-based conclusions are drawn. Recommendations are built on these conclusions, giving appropriate courses of action to address the issues identified in the initial analysis requirements (NATO JALLC, 2016). For our methodology to work, we rely on the Norwegian Armed Forces to identify and communicate the most important issues that need investigating. This requires a structured process for gathering and analyzing experiences made at all levels of the organization – a *Lessons Learned Process* (LLP) (NATO JALLC, 2022). Although it is the responsibility of the Norwegian Armed Forces to establish and maintain a Lessons Learned Capability, because it is critical to our methodology, FFI takes an active role in assisting the Norwegian Armed Forces in managing the LLP. This aid can range from technical solutions to advice on procedures and organization. In this way, we can base our concept development methodology on inputs from the LLP and thereby ensure that we are investigating the correct issues on behalf of the Norwegian Armed Forces.

SUCCESSFUL USE CASE

The three main elements of the concept development process described in the previous section; wargaming, simulation, and live exercises, have been important elements in a successful use case for concept development at FFI that involved a coordinated anti-surface warfare (ASuW) missile attack from forces in the air and maritime domains. This is a joint operation, but the concept development methodology is also highly relevant for MDO.

Figure 2 shows an illustration of a relative timeline for the concept development process for the ASuW missile attack procedure. The timeline starts at month zero with a wargaming phase. In this phase relevant military officers from the air and maritime domains reviewed the existing procedure for a coordinated ASuW missile engagement and continued with refining the procedure together with relevant scientists from FFI. The activities in the wargaming phase included several meetings and discussions, where the concept was further developed. Some of the meetings had only military participants, and others were hosted or attended by scientists from FFI. The meetings were conducted as structured discussions of plans and procedures of an ASuW vignette and can be classified as seminar-type, discussion-based analytical wargames. A crucial output of the meetings was a clearer definition of the research problem, the C2-processes, as well as the military entities and roles involved.

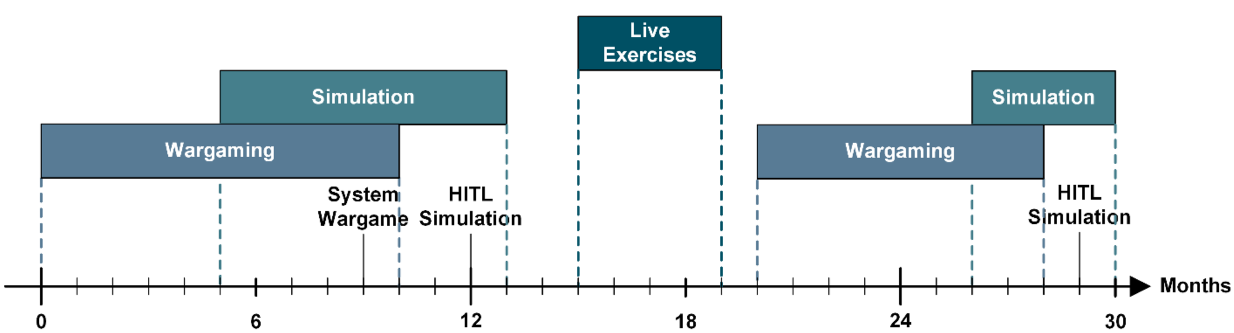

Figure 2. Timeline of Concept Development for ASuW Procedure.

It is difficult to set a specific duration of this phase, but the wargaming activities took place over a period of 1–2 years with increased intensity the last 9–12 months before a HITL simulation experiment was conducted. Parallel to the wargaming phase, the simulation team at FFI began to design and develop a simulation environment in which the concept for the ASuW missile attack procedure could be tested. This is illustrated in Figure 2 by the first simulation phase lasting from month 5 to 13. When the concept had matured into a detailed enough procedure, and the simulation environment was developed to include the ASuW vignette, FFI hosted a type of simulation-based system wargame where relevant officers, from the operational and tactical levels, described how they would conduct their tasks in the joint operation. This event is illustrated in Figure 2 as "System Wargame" at month 9. In this wargame, the simulation environment only served as a computer-based map with military symbols representing combat elements in the ASuW vignette. This activity gave important insight into the plans and procedures of the ASuW missile attack and requirements for the synthetic environment. Further preparations and development of both the synthetic environment and the procedure for the coordinated missile engagement were done before the simulation activity culminated in a HITL simulation experiment executed in FFI's battle lab at month 12 in Figure 2.

In this HITL simulation experiment officers from the operational and tactical levels participated as players on the Blue side and conducted a coordinated missile attack against a surface threat in the synthetic environment. The synthetic environment included computer generated forces (CGF) on the Red side, and the Blue forces were partially automated and partially SAF controlled by the military players. The simulation tools were connected through HLA and DIS (with a gateway to HLA), and the military players used C2 systems that are in operational use as their main user interface. The C2 systems enabled communication between the players, together with chat and voice communication over IP. The HITL simulation experiment lasted two days, and the ASuW vignette was simulated once each day. Minor adjustments to the plan were made between Day 1 and Day 2. During each run, some of the players had access to closed-loop simulation tools for assistance in the decision-making processes. Closed-loop simulations were also used during the AAR for relevant military participants to provide a deeper insight into the area of validity of their actions.

This HITL experiment gave important insight into an experimental procedure for coordinated missile engagement, but simplifications made in the synthetic environment left some questions open for further investigation. Hence, insights, knowledge, and further questions from the HITL simulation experiment contributed as input to several live exercises where different parts of the experimental procedure could be tested further. This phase is illustrated in Figure 2 as "Live Exercises" from month 15 to 19. Testing elements of the operational procedure in a live environment with live systems served as an exam for the concept and gave valuable insights into the procedure, which in turn led to further refinement of the experimental procedure. The set of experiments also highlighted the C2 dependencies of the procedure, and the Air Force and the Navy wanted to explore how these dependencies could be varied.

After a period of analyzing and evaluating the experiments and live exercises, the experimental procedure was once again refined to incorporate alternative C2 dependencies through wargaming. This activity is illustrated in Figure 2 by the wargaming phase lasting from month 20 to 28. In this phase, military officers from the Air Force and the Navy met with FFI in a set of structured discussions to further refine the experimental procedure.

In parallel to this wargaming activity, FFI continued to develop the simulation environment based on experiences from the last HITL experiment, and refined which research questions were relevant to investigate in a new HITL simulation experiment. About 12 months after the live exercises, the refined experimental procedure was again tested in a HITL simulation experiment. This is illustrated in Figure 2 as "HITL Simulation" at month 29. The second HITL simulation experiment was conducted in one day with two runs of the ASuW vignette. In the first run, the original C2 relationship from the previous HITL simulation experiment was used. In the second run, an alternative C2 relationship was tested, providing important insight into the strengths and weaknesses of the experimental procedure. Based on qualitative and quantitative results from the set of meetings, experiments, and exercises in the last 2.5 years, further adjustments and refinements of the experimental procedure were made, and it is now up for evaluation as an operational doctrine.

Although this use case is categorized as a joint operation, we believe that the same method is suitable for concept development for MDO. All phases in the concept development process have their strengths and weaknesses for concept development, but together these tools complement each other and provide as complete an understanding as possible of the new concept.

DISCUSSION

In our concept development methodology, illustrated in Figure 1, the wargaming phase and the HITL simulation experiments can be categorized as discovery experiments. Both wargaming and HITL simulations are typically used to explore new concepts in an open-ended way. They are human-driven processes, relying heavily on human creativity and judgment, and are well suited for gaining insight, uncovering unexpected outcomes, identifying key variables, and generating hypotheses. The closed-loop simulations can be categorized as hypothesis-testing experiments. These simulations can be repeated hundreds or thousands of times to get a statistical distribution of the outcomes, and enable repeatable, measurable testing of specific variables in a controlled environment that is well suited for hypothesis testing. The live exercises phase can be categorized as concept validation. Live exercises provide the most realistic operational context and are used to validate that the concepts actually work in the real world. This phase is used to confirm whether the concept is feasible, effective, and ready for implementation in the real world. As we move through the concept development process, the experiments and activities gradually become more rigid, so development flexibility and room for creativity will gradually decrease. Moreover, the concept will gradually become more detailed and more mature. The categorization of the phases and activities of the concept development methodology, and the characterization of the concept development progress, is summarized in Figure 3.

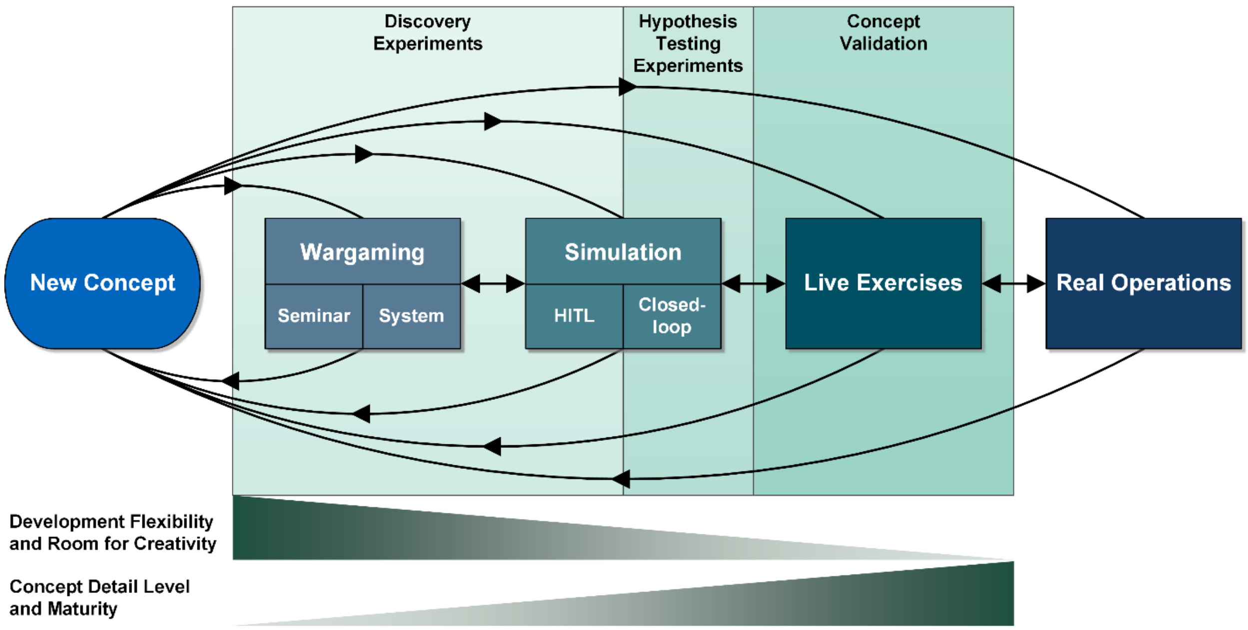

Figure 3. Summary of the Concept Development Methodology with Categorization of the Phases and Characterization of the Concept Development Progress.

Challenges

For our HITL simulation experiments, the technical architecture of our synthetic environment tends to get quite complex, with different simulation tools, command and control information systems (C2IS), in-house-developed user interfaces (UIs), communication tools, and gateways. This complexity makes maintenance, further development, and testing of the synthetic environment challenging and resource- and time-consuming. It is therefore important for us to try to reduce the number of components in the synthetic environment as much as possible.

One challenge regarding the simulation of MDO is that very few, if any, of the simulation tools currently available can represent combat elements, capabilities, and effects across all operational domains at a sufficient and balanced level of fidelity throughout the combat model. In Evensen et al. (2024), we have outlined a set of overall requirements for a synthetic environment for simulation of the future operating environment and MDO, for concept development, experimentation, and analysis. Our preference is to have one simulation system that supports all domains, instead of connecting two or more simulation systems that only support a subset of the five domains. However, a single simulation system that meets all our requirements may not yet exist. We are still in the process of testing promising candidates for a holistic synthetic environment as a basis for experimenting with, and analyzing, combat across all operational domains in an operating environment five to twenty years into the future.

Another challenge is to realistically represent the opposing side in our experiments, especially since we often look at scenarios some years into the future.

We anticipate that conducting live exercises for MDO will be more complex and challenging than for today's joint operations. As a result, it will probably not be feasible to represent all combat elements and capabilities in a purely live environment. An LVC-based environment will therefore be critical for enabling realistic and scalable MDO training and exercises.

Suggested Improvements

To date, our use of closed-loop simulations for hypothesis testing within the concept development methodology has been limited. Further work is needed to identify best practices for applying closed-loop simulations to realistically represent complex operations such as MDO.

Our current concept development process can be characterized as *data-driven*, with decisions made based on data in a largely linear and sequential manner. Our goal is to evolve toward a *data-centric* approach, where the data itself is treated as a core, independent asset. This shift requires a more deliberate strategy for how data is collected, structured, and analyzed at each stage of the process, enabling more effective management, access, and reuse. A data-centric approach will also better support the integration of AI and advanced analytics to enhance experimentation and accelerate insight generation.

Norway's ambitions regarding MDO are quite high, and our concept development methodology needs to become more efficient, for example by accelerating different phases and running more concept development processes in parallel. The use of AI has the potential to make our concept development processes more efficient.

Potential for use of Artificial Intelligence

There is significant potential for the use of AI, particularly generative AI such as large language models (LLMs), to enhance the speed, efficiency, and quality of the concept development process. LLMs can serve as cognitive aids across all phases of experimentation, from early discovery experimentation through to final validation. LLMs can also contribute significantly to documentation, report generation, and knowledge management (Lacks et al., 2025). This includes drafting vignettes, experiment designs, planning documents, and final reports.

In the discovery phase, AI can assist by generating plausible vignettes, behavior for the opposing forces (Chance et al., 2025), and alternative concept variations based on doctrine, operational context, or identified capability gaps. This can speed up preparation for seminar or system wargames and support structured analysis by providing suggested lines of inquiry and essential questions aligned with wargame objectives (Lacks et al., 2025). Additionally, by using automated transcription of wargame discussions, AI can support the analysis by extracting key insights more efficiently.

During the simulation phases, both HITL and closed-loop, AI can assist in proposing parameters to configure and calibrate simulations, translating operational plans into specific tasks executable by CGF, and formulating clear hypotheses based on earlier insights. AI can also support AAR sessions by extracting key findings from transcripts, communication logs, or observer notes. In closed-loop simulations, AI can help synthesize results into concise summaries and suggest next-step experiments to explore the robustness of concepts under varied conditions.

In live exercises, AI may assist experiment teams by recommending logging priorities and supporting post-exercise analysis and evaluation. When used responsibly, AI can help relate observed outcomes back to earlier simulation phases, strengthening the traceability throughout the iterative concept development process.

Overall, AI can facilitate a shift toward a more data-centric concept development process. AI tools can extract structured insights from unstructured data sources such as wargame discussions, simulation logs, and AAR transcripts, enabling better traceability of decision points, supporting cross-phase learning, and accelerating the LLP. As concept development becomes increasingly complex and iterative, AI can act as a cognitive assistant by tagging, summarizing, and linking data across phases and thereby improving the speed, analytical depth, and coherence of experimentation.

While promising, the use of AI must be approached with care, and output should always be reviewed by human SMEs. The value of AI lies in augmenting, not replacing, human expertise, especially in complex and critical areas such as defense experimentation. Incorporating AI also introduces considerations regarding ethical use, operational trust, and the secure handling of sensitive and classified data (Ahmadi et al., 2025).

SUMMARY AND CONCLUSION

In this paper, we have described a structured methodology for enabling MDO through a coherent, iterative concept development process of wargaming, simulation (HITL and closed-loop simulations), and live exercises. The process starts with discovery experiments through wargaming and HITL simulations, continues with hypothesis-testing experiments through closed-loop simulations, and culminates with concept validation through live exercises. As we advance through these phases, concepts gradually become more detailed and mature, until they become operationally ready.

Moreover, we have demonstrated the methodology's effectiveness through a joint-level use case involving a coordinated ASuW missile attack by forces in the air and maritime domains. This example underscores how a structured, iterative approach can generate actionable insights and guide future operational doctrine development.

Importantly, this structured methodology can directly support Norway's and NATO's ongoing shift toward MDO, providing a systematic approach to rapidly refine and validate operational concepts. Going forward, adopting more data-centric processes and integrating AI will further enhance our ability to capture, analyze, and leverage experimental data. Ultimately, this approach not only accelerates concept maturity, but also ensures robust, validated capabilities that maintain operational relevance in increasingly complex and contested environments.

REFERENCES

Ahmadi, E., Green, C., Russell, K., Marx, W., Hill, T., Smith, J., Yohe, M., & Easterling, D. (2025). Are LLMs Too Smart for Their Own Good? *Proceedings of the Interservice/Industry Training, Simulation and Education Conference (I/ITSEC) 2024*, Paper No. 24267.

Alberts, D.S. & Hayes, R.E. (2002). *Code of Best Practice for Experimentation*, The Command and Control Research Program (CCRP) Publication Series.

Appleget, J., Burks, R., & Cameron, F. (2020). *The Craft of Wargaming: A Detailed Planning Guide for Defense Planners and Analysts*, Naval Institute Press.

Burns, S. (edited by) (2015). *War Gamer's Handbook: A Guide for Professional War Gamers*, Newport: United States Naval War College.

Caffrey Jr., M.B. (2019). *On Wargaming: How Wargames Have Shaped History and How They May Shape the Future*, Naval War College Press.

Chance, G., Holland, R., & Halliday, C. (2025). AI for Simulated Wargaming: Intelligent Agents for Multi-Domain Decision & Training Support, *Proceedings of the International Training Technology Exhibition and Conference (IT^{2}EC) 2025*, Technical Paper.

Evensen, P.-I., Martinussen, S.E., Halsør, M., & Bentsen, D.H. (2019). Wargaming Evolved: Methodology and Best Practices for Simulation-Supported Wargaming, *Proceedings of the Interservice/Industry Training, Simulation and Education Conference (I/ITSEC) 2019*, Paper No. 19182.

Evensen, P.-I., Martinussen, S.E., Halsør, M., & Bentsen, D.H. (2021). *Simulation-supported wargaming for assessing force structures – methodology and best practices*, Norwegian Defence Research Establishment (FFI), FFI report 20/02595.

Evensen, P.-I., Halsør, M., & Bentsen, D.H. (2022a). Estimating Relative Combat Effectiveness Using Simulations. *Proceedings of the Interservice/Industry Training, Simulation and Education Conference (I/ITSEC) 2022*, Paper No. 22137.

Evensen, P.-I., Halsør, M., Hoppe, U.-P., & Bentsen, D.H. (2022b). *Measuring combat effectiveness*. Norwegian Defence Research Establishment (FFI), FFI report 21/02310.

Evensen, P.-I., Martinussen, S.E., Halsør, M., & Bentsen, D.H. (2022c). Simulation-Supported Wargaming for Assessing Force Structures. *Scandinavian Journal of Military Studies*, 5(1), pp. 323–333. DOI: https://doi.org/10.31374/sjms.128.

Evensen, P.-I., Hvinden, E.S., Holhjem, H.R., Tveit, D.M., & Eikås, K.D.R. (2024). Requirements for Simulation of the Future Operating Environment and Multi-Domain Operations, *Proceedings of the Interservice/Industry Training, Simulation and Education Conference (I/ITSEC) 2024*, Paper No. 24139.

Hoff, E.Ø., Evensen, P.-I., Holhjem, H.R., Øyan, I.B., & Nygård, H.K. (2012). Simulation in Support of Army Structure Analysis. *Proceedings of the Interservice/Industry Training, Simulation and Education Conference (I/ITSEC) 2012*, Paper No. 12088.

Hoff, E.Ø., Evensen, P.-I., Holhjem, H.R., Øyan, I.B., & Nygård, H.K. (2013). Interactive Simulation to Support the Transition of Forces. *Proceedings of the NATO Modelling and Simulation Group (NMSG) Annual Symposium 2013 (STO-MP-MSG-111)*, Paper No. 6.

Kass, R.A. (2006). *The Logic of Warfighting Experiments*, The Command and Control Research Program (CCRP) Publication Series.

Lacks, D., Kackley, T., & Godfrey, G. (2025). Applying AI Techniques to Improve Wargaming Design and Delivery, *Proceedings of the International Training Technology Exhibition and Conference (IT^2EC) 2025*, Technical Paper.

Martinussen, S.E., Tansem, I., Evensen, P.-I., Rødal, H., & Bore, I. (2008). Simulation in Support of Defence Structure Analysis. *Proceedings of the NATO Modelling and Simulation Group (NMSG) Annual Symposium 2008 (RTO-MP-MSG-060)*, Paper No 22.

North Atlantic Treaty Organization (NATO) Allied Command Transformation (ACT). (2021). *NATO CD&E Handbook – A Concept Developer's Toolbox*. Version 2.10.

North Atlantic Treaty Organization (NATO) Joint Analysis and Lessons Learned Centre (JALLC). (2016). *Joint Analysis Handbook*. 4th Edition.

North Atlantic Treaty Organization (NATO) Joint Analysis and Lessons Learned Centre (JALLC). (2022). *The NATO Lessons Learned Handbook*. 4th Edition.

North Atlantic Treaty Organization (NATO) Standardization Office (NSO). (n.d.). NATOTerm – The Official NATO Terminology Database. Retrieved April 28, 2025, from https://nso.nato.int/natoterm.

North Atlantic Treaty Organization (NATO) Standardization Office (NSO). (2022). *Allied Joint Doctrine*. Allied Joint Publication (AJP)-01, Edition F, Version 1.

Norwegian Defence Research Establishment (FFI) (2024). Research for our future defence – FFI's 20-year strategy for 2044.

Norwegian Ministry of Defence (Norwegian MoD) (2024). Prop. 87 S (2023–2024) – The Norwegian Defence Pledge: Long-term Defence Plan 2025–2036.

Numrich, S.K., Woods, K., Resnick, J., & Jackson, J. (2017). Innovation through Discovery Experimentation. *Proceedings of the Interservice/Industry Training, Simulation and Education Conference (I/ITSEC) 2017*, Paper No. 17317.

Numrich, S.K. & Woods, K. (2024). Exploiting Experimentation: A Managerial Challenge. *Proceedings of the Interservice/Industry Training, Simulation and Education Conference (I/ITSEC) 2024*, Paper No. 24151.

Perla, P.P. (1990). *The Art of Wargaming: A Guide for Professionals and Hobbyists*, Naval Institute Press.

Rubel, R.C. (2006). The Epistemology of War Gaming, *Naval War College Review*, Volume 59, No. 2, Article 8.

The Technical Cooperation Program (TTCP). (2006). *Guide for understanding and implementing defense experimentation (GUIDEx)*. Version 1.1.

Willis, C., Bayer, J., Kelly, J., & Anderson, S. (2023). A Hybrid Approach to Combat Simulation Experimentation, *Proceedings of the Interservice/Industry Training, Simulation and Education Conference (I/ITSEC) 2023*, Paper No. 23106.