Improving Training and Education Supply Chains by Harnessing Data Pipeline Observers

Ms. Erica Dretzka, Mr. Jordan Gottlieb
Office of the DoD Chief Digital and AI Officer
Washington, DC
erica.l.dretzka.civ@mail.mil;
jordan.gottlieb.ctr@mail.mil

Mr. Brent Smith
ADL Initiative (SETA)
Orlando, FL
Brenton.t.smith2.ctr@mail.mil

ABSTRACT

Governments, corporations, and consumers embattled by relentless cyberattacks have a growing imperative for secure supply chain assurance. Current industry standards have proven insufficient in mitigating supply chain risks from compromise and attacks. Many critical systems (cloud software, weapon systems, expert systems, human, and modeling/simulation) rely on extensive networks of interconnected capabilities. These interconnected systems encompass sensitive processes and data used for configuration, operations, and training.

Following Executive Order 10428, Improving the Nation's Cybersecurity, DoD has widely adopted Software Bills of Materials (BOMs) to catalog the range of different components, libraries, interfaces, and specifications that each application is built from. Building on SBOM adoption, the [eXtensible] BOM documents all aspects of the training and education supply chain across each platform's hardware, software, data, and human components required to operate and maintain the software lifecycle. Broad applicability across various training and education platforms and delivery modalities requires adaptable metadata patterns and structures that describe each component while maintaining provenance across dozens of layers of sub-dependencies.

Across these various components, unauthorized data manipulation for AI algorithms has serious consequences. Corrupt data compromises the integrity of AI models and leads to inaccurate predictions, biased outcomes, and detrimental decisions. Physical security relies on defense in depth, where exterior cameras detect perimeter motion, keyed locks protect the front door, elevators control floor access via a badge, and a hand palm reader protects the research laboratory door. Layered cybersecurity architectures follow similar principles in protecting various aspects of DoD networks. However, additional security techniques are necessary to observe the myriads of atomic data manipulation and fundamental zero trust principles. This paper documents the processes being used to create and maintain [x]BOMs to support DoD's Enterprise Digital Learning Modernization (EDLM) reform. EDLM [x]BOMs are being developed to catalog the data, inputs, commands, environmental characteristics, and outputs of each software component, including cryptographic hashing, prior to inclusion in an [x]BOM fabric to establish provenance with high-confidence characteristics.

ABOUT THE AUTHORS

Erica Dretzka: Ms. Erica Dretzka is a seasoned data scientist with over 20 years of experience in various industries, including Insurance, Energy, and National Defense. She has established two data science teams inside the Department of Defense (DOD) and led the development of advanced Artificial Intelligence (AI) and Machine Learning (ML) models. She focuses on employing engineering-based methods to design the optimal reference architecture and bridge strategy to support AI and data-backed mission support at the scale and resilience required for DOD.

Jordan Gottlieb Mr. Gottlieb is passionate about applying innovation and collaborative techniques for strategic synergies to unify stakeholders across an enterprise and create a data interoperability model across all of industry. He is currently focused on defining and implementing a true data mesh as well as defining an actualized Zero Trust model that is complimentary to current cybersecurity capabilities. His broad background includes system engineering, program management, strategic planning, business operations, government acquisition, and executive support.

Brent Smith: Mr. Brent Smith is a Software Systems Architect with over 25 years of experience designing and developing learning technologies for government stakeholders, defining an enterprise L&D data strategy to meet DoD-wide objectives, and establishing chains of research that align with strategic goals of enabling a data-drive ecosystem. As the ADL Initiative R&D Principal, Mr. Smith helps ensure the ADL Initiative research plan is aligned with its overall strategy.

Improving Training and Education Supply Chains by Harnessing Data Pipeline Observers

Ms. Erica Dretzka, Mr. Jordan Gottlieb (SETA)
Office of the DoD Chief Digital and AI Officer
Washington, DC
erica.l.dretzka.civ@mail.mil; jordan.gottlieb.ctr@mail.mil

Mr. Brent Smith
ADL Initiative (SETA)
Orlando, FL
Brenton.t.smith2.ctr@mail.mil

INTRODUCTION

The Services have long recognized the power of data and its ability to fuel automation, reduce workload, streamline processes, enable fresh insights, and achieve never-before-seen efficiencies. From optimizing individual learning pathways to enhancing organizational capabilities and informing institutional curriculum improvements, data establishes a foundation for a more agile, efficient, and future-ready DoD workforce. Across various DoD organizations, the interconnectedness of competencies, learning resources, and performance data promotes a more agile workforce capable of swift adaptation to technological advancements and mission demands. The Enterprise Digital Learning Modernization (EDLM) program and infrastructure establishes a cohesive link between training outcomes, career field competencies and credentials detailing necessary knowledge, skills, abilities, and tasks (KSATs) that are developed by DoD functional communities.

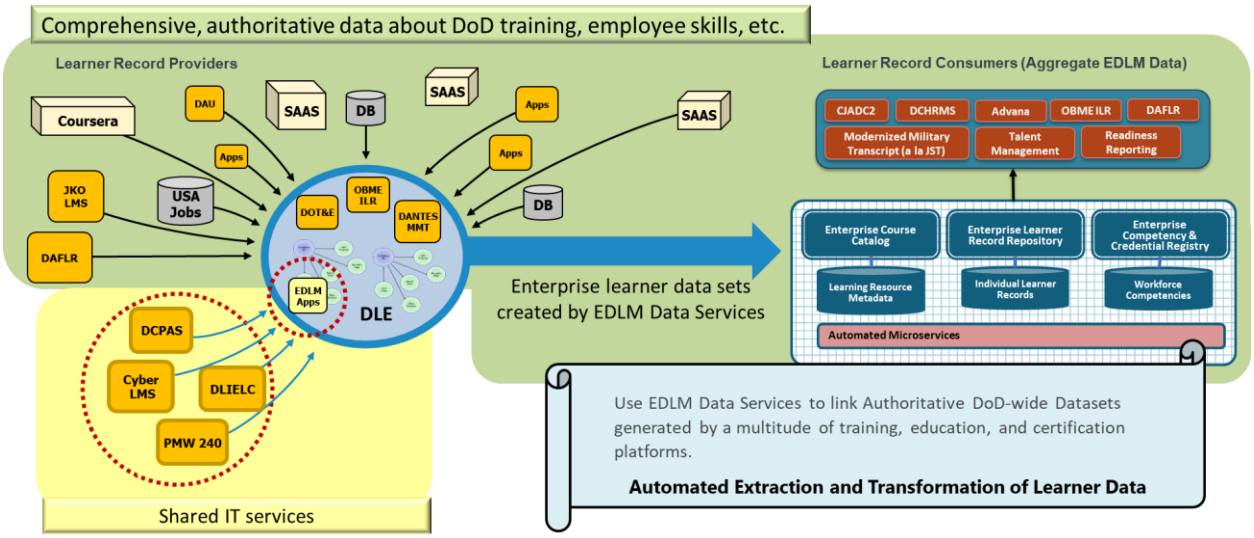

Figure 1. Enterprise Digital Learning Modernization (EDLM) Data Services

With this power comes great responsibility. As shown in Figure 1, learner data also enables valuable insights into Readiness Reporting, Combined Joint All-Domains Command and Control (CJADC2), and cross-Service Talent Management initiatives. Throughout the continuum of career-long learning in DoD, a multitude of different Information Technologies work together to Analyze, Design, Develop, Implement, and Evaluate different education, training, or certification programs across the department. Data privacy and security measures must be applied uniformly across the lifecycle of learner data from the point of delivery and across the multitude of connected systems. When combined, the detailed learner data generated by these systems enables organizations to track and validate the acquisition of specific skills and competencies, providing both educators and employers with a clear picture of an individual's readiness for the workforce.

Industry 4.0 and Industry 5.0 represent transformative waves in the manufacturing and industrial sectors, driven by technological advancements. Industry 4.0, also known as the Fourth Industrial Revolution, introduced the integration of automation and data exchange in manufacturing technologies, including cyber-physical systems, the Internet of Things (IoT), and cloud computing [European Commission, 2021]. It marked a shift towards "smart" factories capable

of more efficient and customizable production processes. Industry 5.0 builds upon this foundation by emphasizing the collaboration between humans and machines, aiming to create sustainable and resilient industries that are more human-centric [SAP, 2024]. This approach not only seeks to enhance productivity but also to ensure that technological progress contributes positively to societal and environmental goals.

The impact of these initiatives is far-reaching, affecting workers, businesses, and economies worldwide. For the DoD, the implications are significant in terms of training and education. As Industry 4.0 and 5.0 technologies evolve, there's a growing need for a workforce skilled in digital, technical, and collaborative competencies. The DoD's EDLM program is working to establish the data infrastructure required to adapt to these changes by enabling the development curricula that address the demand for new knowledge and skills, ensuring that military and civilian personnel are prepared for the future of work in an increasingly automated and interconnected world. This is crucial for maintaining operational effectiveness and national security in an era where technological superiority is paramount.

EDLM Use Case:

In 2021, DoD leadership introduced several initiatives to prioritize DoD-wide talent management. These efforts aimed to address workforce development, personnel policy, and people management by elevating the existing DoD functional community governance framework [DoD Instruction 1400.25]. The 28 DoD Functional Communities (FC) play a crucial role in DoD-wide workforce planning, competency assessment, and identifying mission-critical gaps. Each DoD FC is building out labor frameworks that include workforce competency models to assess workforce KSATs, build career roadmaps to ensure employees have a clear 'line of sight' for career development, and individual competency development plans to help their workforce progress and meet professional standards.

Each functional community utilizes a wide range of different Information Technologies, tools, and data formats both internal and external to DoD to describe the competencies, credentials, assessments, and learning resources required to support their workforce elements, work roles, and KSATs. A 2020 study estimated that there are about 250 types of cybersecurity frameworks and standards in use globally throughout the world. In many instances, industry utilizes multiple cybersecurity frameworks and standards to suit their needs. Today, many companies use more than one framework and standard in their business operations [Syafrizal, 2020]. EDLM's adoption of [x]BOMs helps provide a comprehensive inventory of all training and education applications and their technology stacks to improve Supply Chain transparency across connected systems. As shown in Figure 2, an [x]BOMs is used to decompose the cyber range into its atomic services and the required resources for each platform's hardware, software, data, and human components required to operate and maintain the software lifecycle.

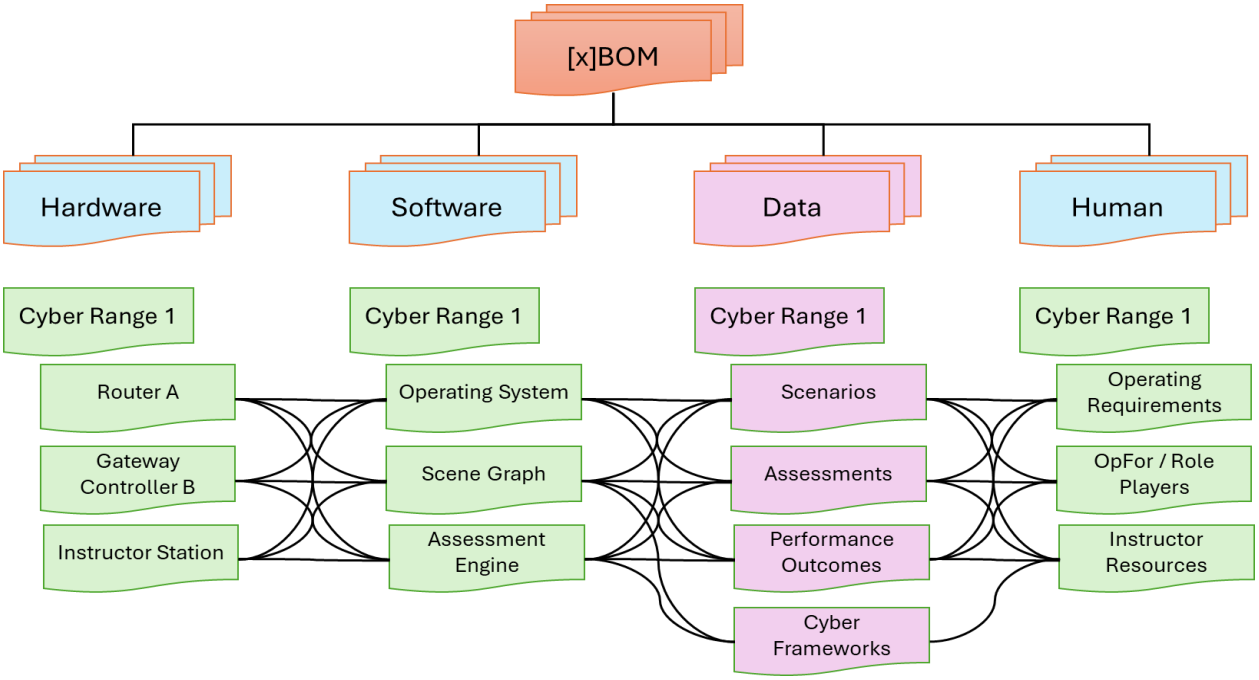

Figure 2. [x]BOMs describe the building blocks of any application

The [x]BOM serves as a comprehensive framework for the EDLM data services to connect, share, and interpret learner data being generated by DoD's training and education community. They're designed to standardize and structure different types of learner data, facilitating semantic interoperability across various DoD Functional Communities and the Services. This interoperability is crucial for ensuring that learner data can be seamlessly exchanged, understood, and utilized across different systems and organizations.

To enable successful training, the data from each platform must be captured and aligned with the relevant cybersecurity frameworks used within the jobs, work roles, and position descriptions used by the operational force. By defining the nuances of each organization and each digital learning platform, [x]BOMs enable the precise interpretation of learner data at the atomic level. This means that every piece of data, no matter how small, is tagged and categorized in such a way that its meaning and context are preserved and understood universally within the DoD. As a result, [x]BOMs empower the DoD to track individual learning progress, tailor educational experiences, and optimize training resources effectively, ensuring that personnel are equipped with the necessary skills and knowledge to meet the evolving demands of their roles.

More recently, EDLM is leveraging the concept of *"Data Pipeline Observers"* to continuously monitor the data generated by the different systems to detect anomalies in expected patterns that might impact data integrity or data security. Data pipeline observers are a critical component in the management of EDLM's data ecosystems, particularly in ensuring the quality and integrity of data as it flows from source to destination. These observers act as watchdogs, continuously scanning the data pipeline and comparing with the [x]BOM to validate data consistency, accuracy, and completeness. They are designed to detect anomalies, outliers, or patterns that deviate from established norms, which could indicate potential issues with the data. This enables an automated layer of protection that can identify and flag errors in real-time, preventing the propagation of corrupt data through the system. This is especially important in large-scale data operations where manual monitoring is impractical.

In the context of the using learner data for decision making where data integrity is paramount, data pipeline observers are a vital tool that helps ensure the data used in decision-making processes is reliable and that operational security is maintained by safeguarding against data breaches and cyber threats. By integrating these observers into the EDLM data management strategy, the ADL can enhance the overall security posture and efficiency of EDLM data services. The combination of [x]BOMs, Data Pipeline Observers and the Chief Digital and Artificial Intelligence Office's Data Mesh Reference Architecture (DoD CDAO, 2024) enables EDLM data services to implement policy-based access across the lifecycle of each piece of the learner data Supply Chain.

But what really is an [x]BOM?

An [x]BOM is a foundational concept within the DoD that represents a strategic approach to managing and integrating vast amounts of data across the organization. Key to utilizing [x]BOMs across the department is a critical infrastructure component designed to enable DoD-wide insights and decision-making.

[x]BOMs rely on Canonically Controlled Vocabularies to ensure that terminology is standardized at the atomic level across different branches and units. DoD-wide Unique Identifiers provide a consistent way to reference and track all items in an [x]BOM and ultimately each data element through its different uses across the DoD's complex network of decision making. Microservices play an important role tying these capabilities together to connect legacy software systems to other enterprise systems. These small, self-contained programs interact through well-defined interfaces and protocols, allowing for the seamless integration of new technologies with older systems.

Across DoD, an [x]BOM is not constrained to a prescribed list of components. The term extensible refers to the fact that using controlled vocabularies and unique identifiers, each [x]BOM can be extended by different user communities across the department. For the purposes of this paper, the EDLM use case is constrained to the following components.

- Software: This covers the applications and operating systems that facilitate the delivery of training content, data management, and the overall functionality of the platform. Software BOMs are broadly documented in DoD with multiple standards in place, including Linux Foundation's SPDX and OWASP's CycloneDX. The [x]BOM infrastructure allows DoD to create a fabric that enables all formats by promoting interoperability at the atomic level.
- Hardware: These are the physical devices such as servers, computers, and networking equipment that support the infrastructure of the training environment. The United States Cybersecurity and Information Security Agency's Hardware BOM Framework (CISA, 2023) provides consistent naming conventions for

attributes of hardware components and provides a format for identifying and providing information about the different types of components used in a training or education platform. Each Hardware BOM provides important information about each components Supply Chain including important information about maintenance, replacement parts, and logistics.

- Data: A critical element, data comprises the curriculum, learner records, performance metrics, and other educational materials that are essential for the training process. Within EDLM, each training and education platform's data assets include a unique family of subclasses that align to the different types of data supported by EDLM data services. Software systems in use across the Human Capital Supply chain are interdependent. Different learning experiences are coupled with learning outcomes which are ultimately linked to jobs and assignments. AI/ML components typically have their own supply chains but are currently being integrated into the Data BOMs developed for EDLM. Different commercial solutions include Google Data Cards, Google AI Cards, SPDX AI BOMs, and Hugging Face Model Cards.
- Manpower & Resources: This refers to the instructors, administrators, and learners who interact with the system. Their roles, responsibilities, interactions, and other required resources are vital for the efficacy of the training programs. Different training platforms are used to support different delivery modalities which require human and non-human resources to support their operation. Delineation of these roles and resources is important to securing how data is secured at the point of delivery. Manpower BOMs require a clear strategy to describe the collaboration between human workers and machines to optimize training, education, and certification processes and enable a culture of continuous improvement.

Together, these elements form the backbone of the EDLM [x]BOM framework, which can deliver comprehensive, real-time data that enables insights that are essential for the DoD's operational success. As shown in Figure 3, different institutions / organizations use a combination of different tools and technologies to manage the design, development, and delivery of instruction to their students. These applications are deployed in concert with different training management systems, student information systems, learning management systems, assessment tracking / certification tools, and exercise control systems. All the different systems within an organization, and their components / subcomponents, must work in harmony to ensure that the system is effective, secure, and capable of adapting to the evolving needs of the DoD. The [x]BOM serves as a blueprint, detailing how these elements interconnect and operate within the larger ecosystem of DoD's educational initiatives.

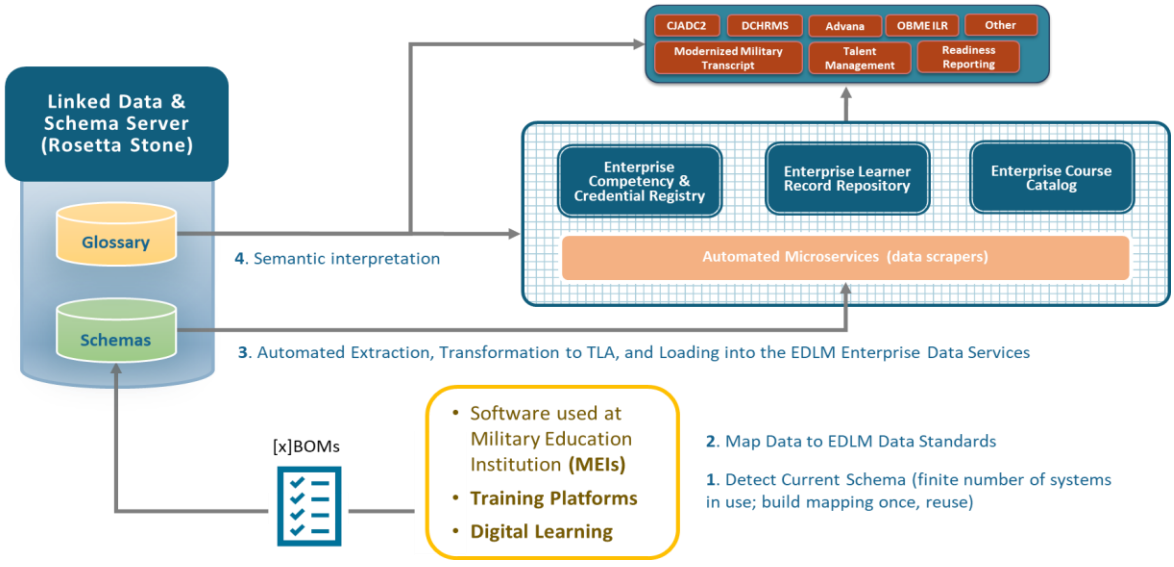

Figure 3. EDLM Data Services leverage [x]BOMs to connect Training and Education Systems with other HRM applications

[x]BOMs play a pivotal role in enhancing the provenance, security, and integrity of EDLM data assets by providing a detailed inventory of data assets, including their origins, patterns, profiles, handling, and modifications, [x]BOMs ensure a transparent trail of data lineage that is crucial for provenance and supports the verification of data sources and transformations throughout its lifecycle using data pipeline observers.

[X]BOM IMPLEMENTATION STRATEGIES IN TRAINING AND SIMULATION

Military Education Institutions and Department of Defense (DoD) Training commands utilize a diverse array of learning tools, technologies, and platforms to deliver comprehensive training and education to service members. These range from traditional classroom settings equipped with multimedia presentations to advanced virtual reality simulations and online learning management systems. Each tool and platform is tailored to meet specific training objectives, whether it's for combat readiness, technical skills acquisition, or leadership development. However, learning objectives, outcomes, and competencies are often defined differently across functional communities and the Services, reflecting the unique requirements and missions of each group. This variance can pose challenges to interoperability and data integration.

The introduction of [x]BOMs has potential to revolutionize the way these educational resources are managed and deployed. An [x]BOM, once created for a particular platform, serves as a detailed template that encapsulates all the necessary components and configurations. This template can then be replicated across the department for every instance of the system, ensuring consistency and reducing the time and resources required to set up and integrate new instances of the platform. As shown in figure 4, [x]BOMs provide the framework for interoperability at the atomic level by defining the granular details of each training platform, ensuring that learner data is not only consistent and portable but also meaningful across different contexts. This is what promotes interoperability with the broader DoD data mesh.

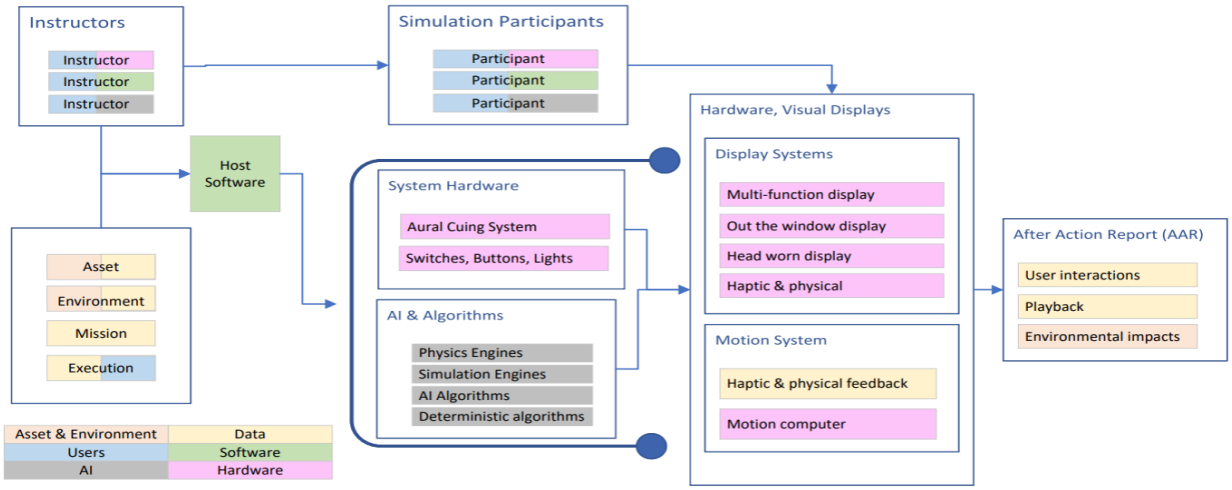

Figure 4. EDLM [x]BOMs connect local, siloed data about each platform to the DoD-wide Data Mesh

Within EDLM, [x]BOMs enable a seamless flow of information between localized systems using siloed data and the enterprise data mesh, allowing for the aggregation and analysis of data to inform decision-making and improve the overall effectiveness of DoD training programs. This level of detail and control is essential for maintaining the integrity of the data generated by each system and ensures that data aligns with EDLM and ultimately DoD manpower requirements.

Training simulators provide a good exemplar on the different types of EDLM [x]BOMS that are required to adequately describe the simulator systems components and their supply chain. [x]BOMs are comprehensive lists that detail every component and item required for the operation and maintenance of a simulator or any training and education platform. A simulator is a complex piece of technology that requires many components to all work together in unity to ensure everything is functioning properly and is creating a realistic and immersive experience for the trainee [AVT Simulation, 2024]. At a rudimentary level, a simulator platform's [x]BOM may be broken down into the following high-level functions.

1. Software: Simulator uses one or more models and software to emulate a real-world experience that stimulates various hardware components of the simulation and displays the resulting environment on large screens for simulation participants and smaller screens for instructors and observers of the simulation scenario. The software accepts mission planning inputs from instructors and mediates exchanges between simulation participants, the system hardware, and interoperates with physics engines and other algorithms to calculate

the impacts of these interactions on the models being used to drive the simulation scenario. Outputs from the simulation are reflected in the simulation hardware and visual displays to drive participant perceptions. User interactions are tracked, analyzed, and stored for potential playback in an After-Action Review. Each simulator platform is typically comprised of multiple software systems that use different languages, frameworks, messaging protocols, and interface specifications. A key component of the Software BOM is that ability to document all building blocks used in the development of the simulator's software components and to identify the interplay with other components within the [x]BOM.

2. Hardware: A user of the simulator interacts with the various software components by using any number of hardware components including input / output devices, displays, motion platforms, image generators, and the myriads of switches, buttons, instruments, alarms, warning light, indicator lights, and levers that might be integrated into the simulation hardware platform. Each of these components has its own supply chain, maintenance procedures, resource dependencies, and risk. The Hardware BOM component of the [x]BOM creates a consistent, repeatable way to name and describe the different hardware components, sub-components, documentation, attributes, and their interplay with other [x]BOM components such as the firmware used by a piece of hardware that is defined in the Software BOM section of the [x]BOM. Hardware BOMs are often structured hierarchically to show dependencies between the simulator's components, assemblies, OEM providers, and other manufacturing details.
3. Data: Data is embedded across all components of a simulation platform from the different models used to drive visuals, behaviors, ballistics, and impacts within each scenario. This type of data is used internally to the simulation platform while other data used in the Analysis, Design, Development, Implementation, and Evaluation of each platform may be used in other platforms to support the lifecycle of each platform. Initial deployment of EDLM data services includes 3 different buckets of data that describe 1) the learning resource or event, 2) the skills and competencies being taught or assessed, 3) learner performance, credentials, and outcomes that a learner obtains after completing the resource. The data exchanged between the software, hardware, and user components may be both quantitative (e.g., numeric) and qualitative (e.g., narrative, audio/video). The Data BOM components of the [x]BOM utilize EDLM data services to extract and transform the data into the EDLM data model which leverages controlled vocabularies built to support the ADL standards described in DoD Instruction 1322.26 for Distributed Learning.
4. Manpower & Resources: Resource BOM components play a vital role in managing supply chain logistics by providing a clear and structured overview of a simulator's resource requirements. They help in tracking the usage rates of consumables, the maintenance schedules of durable equipment, and the availability of personnel. By integrating [x]BOMs with supply chain management systems, organizations can automate procurement processes, schedule timely replenishments, and even predict future needs based on historical data. This results in a streamlined process that minimizes waste, reduces costs, and enhances the overall efficiency of resource management in training and educational environments. For example, an [x]BOM for a medical simulation, might list specific resources such as patient simulators, live role players, and moulage kits, which are used to simulate various medical scenarios and injuries. This level of detail allows for precise planning and allocation of resources, ensuring that each training session can be conducted without delays or shortages of materials.

Each component in the multi-layered ecosystem described above is highly interconnected to the others. However, each component also serves additional purposes when connecting a training or education platform to the enterprise data mesh via EDLM data services. Within the context of EDLM, [x]BOM components are used to connect these applications to the different Human Resource Management systems in use across the Services and 4th estate to recruit, retain, and develop a qualified workforce. Privacy and security of learner data is a primary concern for all training and education platforms. Protections must also be put in place to protect the data from nefarious actors that may attempt to inject malicious data into the system.

Layered Cybersecurity and Zero Trust Architecture

EDLM data services collect data about DoD manpower requirements and the knowledge, skills, activities, and tasks being taught within DoD. More importantly, EDLM learner data provides valuable insights into workforce proficiency levels which can be used to expose gaps and critical workforce training needs. Strong privacy and security protections are essential to protect sensitive information from unauthorized access and potential breaches. Given the diverse ecosystem of platforms, organizations, user roles, and non-person entities that use EDLM data, it is critical that access to different data assets is meticulously managed and aligned with the individual's role, clearance level, and need.

Zero Trust (ZT) is a comprehensive strategy that prescribes the digital protective measures required to protect DoD networks, applications, and data. When implemented, ZT provides a DoD-wide approach for managing identity, credentialing, and access control to physical environments (e.g., server rooms, training ranges) and digital systems (e.g., hardware, software, data). [x]BOMs enhance the ZT management of EDLM data by embedding policy-based permissions into the metadata of each data element. This approach allows for dynamic access control, where permissions can adapt to changes in policy or the environment. As a result, [x]BOMs not only facilitate the secure and compliant use of resources but also contribute to the efficient governance of data within the supply chain, ensuring that each element is handled according to its associated policies at every stage of its existence.

Policy Based Access Control (PBAC) is governed by predefined policies that control who can view or modify each data element. The EDLM data strategy emphasizes the preservation of raw learner data to protect the evidentiary chain of how and why competencies, skills, and credentials have been conferred to a learner. EDLM [x]BOMs are integral to the privacy and security protections required by the different Military Education Institutions, DoD schoolhouses, and DoD training commands. Because EDLM is collecting learner data across industry, academia, and government, data privacy and protection requirements must also meet the Family Educational Rights and Privacy Act (FERPA). FERPA includes strict guidelines on how academic institutions can share data, what data can be shared, and when exceptions for some data types may be made [Department of Education, 2024).

The EDLM program leverages Role Based Access Control (RBAC) to define clear access boundaries based on roles within the DoD, while Attribute Based Access Control (ABAC) allows for dynamic access decisions based on attributes that can change over time, such as the mission's context or data classification changes. RBAC operates on the principle that access to resources is granted based on the roles of individual users within an organization. This ensures that personnel can access only the data and systems necessary for their duties, thereby upholding the 'least privilege' security principle. ABAC provides a more granular level of control by considering multiple attributes, such as the context of access, the sensitivity of the data, and user attributes like clearance level, geographic location, unit ID. Table 1 shows the different actors and their roles that guide how they will be accessing EDLM systems. Each Actor is treated as a unique identity with a description and characteristics.

Table 1. Actors and Roles

Actor	EDLM User Type	Description
DoD Employee (Student)	Personal User	Access is inherently authorized for the individual the data resource pertains to, third party access may be explicitly delegated by the individual or through relationships such as dependency or power of attorney.
Training Manager (Instructor)	Distinct User	Access is automatically authorized based on core attributes. Training managers and instructors have access to data related to competencies, credentials, learning resources, and individual learner data for students they are responsible for. Read/write privileges may be restricted based on organizational policies
Supervisor	Distinct User	Access is automatically authorized based on core attributes. Supervisors have access to data related to competencies, credentials, readiness dashboards, workforce analytics, and relevant learner data. Read/write privileges and access to data may be restricted based on organizational policies
Career Field Manager	Limited User	Access requires explicit authorization based on role, need to know, or specialized attributes. Career Field Managers play a crucial role in overseeing and managing the development and progression of individuals within specific career fields or occupational specialties. They need access to data about the skills, competencies, and developmental needs of their employees.
Workforce Planner	Limited User	Access requires explicit authorization based on role, need to know, or specialized attributes. Workforce planners focus on assessing and ensuring the proficiency of their workforce against current and future requirements. They need access to organization proficiency levels and skill gaps against manpower requirements embodied in labor frameworks that include the knowledge, skills, abilities, and tasks required by each work role
Functional Community Manager	Functional Privileged User	Access includes approval authorities or other operations management capabilities and requires explicit authorization
EDLM Admin	IT privileged user	Access allows management of information technology such as systems, networks, or databases, or management of audit logs, E2E tests, deployment scripts, and other tasks. Requires explicit authorization.

Within EDLM, each actor is assigned a role that aligns with a DHRA user type. The role defines how each actor is authenticated, registered, and logged into each EDLM application. Each Role contains sufficient characterizations to understand authorizations, permissions, privileges, authorities, etc. These characteristics are divorced from the Actor. A variety of attributes are being employed to ensure that only authorized individuals have access to sensitive data. These attributes can include role designation, clearance level, location, time of access, and the specific training module being accessed. By utilizing a combination of these attributes, the DoD can create a robust security posture that dynamically adapts to varying levels of trust and changing conditions. Table 2 includes an abbreviated example to highlight the different attributes that can be used to restrict access to EDLM data.

Table 2. Attributes

Actor	Attribute Type	Description
DoD Employee (Student)	Course Enrolled in	This attribute restricts access to learning materials and IT systems that host courses an individual is enrolled in.
DoD Employee (Student)	Clearance Level	This attribute restricts access to learning materials and IT systems an individual can access.
Training Manager (Instructor)	Assigned Courses	This attribute restricts access to learning materials, learner records, and systems available for the courses they are assigned.
Training Manager (Instructor)	Subject / Domain	This attribute restricts access to learning materials, composite learner data, instructor resources, and skill / competency frameworks to the topics they're responsible for teaching.
Supervisor	Service / Department	This attribute restricts access to only use learner data within their department
Supervisor	Certifications Required	This attribute restricts access to mandatory training / certification requirements for specific staff
Career Field Manager	Career Field(s) responsible for	This attribute restricts access to learning materials, learner records, and IT systems an individual can access based on the career fields an individual is responsible for.
Career Field Manager	Occupational Specialty	This attribute restricts access to learning materials, learner records, and IT systems an individual can access based on the occupational specialties assigned to support that career field.
Career Field Manager	Service / Department	This attribute restricts access to learning materials, learner records, and IT systems an individual can access based on the Service a career field manager supports
Workforce Planner	Functional Community	This attribute restricts access to learning materials, learner records, and IT systems an individual can access based on the Functional Community a workforce planner supports.
Workforce Planner	Service / Department	This attribute restricts access to learning materials, learner records, and IT systems an individual can access based on the Service a workforce planner supports
Workforce Planner	Clearance Level	This attribute restricts access to learning materials, learner records, and IT systems an individual can access.

Together, these controls enable a robust Policy-Based Access Control (PBAC) system that can enforce complex policies over who can access what data, when, and under what conditions. [x]BOMs support these types of controls by creating a detailed inventory of all the hardware, software, resources and associated data elements required for each training and education platform / system being used across the DoD. They can be designed to include role and attribute information for each data element, which then informs the PBAC system. Roles can be cleanly separated from each other, and elevated privileges can be defined by building policies around the different roles and attributes of individuals accessing the data. More critically, a role can be redefined once across the entire enterprise, and all actor privileges are immediately updated with the new role specification.

When encoding roles and attributes into each data element, it is beneficial to separate the targets of ABAC into multiple categories to ascribe characteristics being codified. EDLM learner data is accessed following the set policies throughout its lifecycle from the point of generation and across the various ways that data is used across the department. For EDLM, these are the 4 major components of the EDLM [x]BOM:

- Software Assets: describe roles, permissions, and access restrictions based on EDLM schedules
- Hardware Assets: describe system requirements, roles, operations & maintenance processes, logistics
- Data Assets: Align to EDLM data models, describe roles, permissions, and access restrictions
- Manpower & Resources: Describe staffing requirements, dependencies, roles, and permissions

PIPELINE OBSERVERS

Supply Chain Risk Management (SCRM) is an essential element of a comprehensive protection posture. [x]BOMs can be used to capture the distinct and discrete componentry of any object, atomic or compound, and provides the basis to perform a detailed inspection of the full lineage characteristics for each training and education platform used across the department. Utilizing [x]BOMs to inventory DoD's training and education IT creates the richness of information necessary to gain insight and composable details to understand the origin and evolution of an item throughout its lifetime. By requiring the [x]BOMs to be machine-readable and linking data elements to controlled vocabularies, then the entire SCRM process can be fully automated and dynamic adaptations for a constantly changing environment becomes possible.

This information contained in [x]BOMs creates the basis to perform SCRM analysis and assess the potential risks that any item injects into the environment based on established tolerance thresholds. A data pipeline observer in the context of EDLM is a specialized component designed to monitor and analyze the flow of data between systems. It functions by continuously scanning the data being generated, looking for patterns and anomalies, and ensuring that the data adheres to predefined norms and quality standards. Utilizing [x]BOMs, data pipeline observers understand the expected outputs and resource requirements of the system, which aids in establishing benchmarks for normal operation. Data pipeline observers effectively track whether the educational content, training modules, and learner performance data are being generated and utilized as intended. This is particularly important in complex systems where multiple components work in tandem to deliver educational services. By comparing real-time data against the [x]BOMs standards, data observers can promptly identify discrepancies, flagging issues related to data quality or integrity that could impact the effectiveness of the training and education delivered.

Data pipeline observers are used to validate data quality, consistency, and security, ensuring that only verified and relevant data is utilized for analysis and reporting through EDLM data services. These observers help protect the integrity and reliability of the data fed into other connected DoD systems, artificial intelligence and machine learning systems, and numerous decision support tools used to interpret learner data in terms of Readiness Reporting, Combined Joint All-Domains Command and Control. DoD-wide talent management. Ultimately, data pipeline observers contribute to a robust data governance framework that underpins the success of data-centric initiatives within any organization.

Within EDLM, these goals can be achieved by deploying data observers as side-car containers that run alongside the primary microservice containers, providing a lightweight mechanism to observe the inter-container communication without interfering with the core functionalities. They monitor the data in transit from one system to another within the Kubernetes pods, ensuring that the data integrity is maintained throughout the process. This integration is crucial for maintaining a high level of data quality at the point of educational delivery, as it allows for real-time monitoring and immediate response to any potential issues, thereby safeguarding the reliability and trustworthiness of the EDLM data services.

FUTURE DIRECTION AND RESEARCH OPPORTUNITIES

The EDLM program's adoption of [x]BOMs represents a significant stride towards cataloging the essential hardware, software, data, and resources used or generated by DoD's training and education community. This systematic approach streamlines the management of these systems and their resources and fosters a cohesive environment for data interoperability between legacy training and education platforms with other connected systems across DoD's data mesh. Next steps for EDLM include the establishment of data sharing agreements and authorities to connect to different systems to build [x]BOMS for numerous disparate training and education platforms (e.g., Learning Management Systems, Learning Resource Repositories, Assessment tracking tools, and other software applications that store EDLM data. However, from a DoD-wide perspective, additional work is required to establish globally unique Identifiers and Controlled Canonical Vocabularies to support EDLM interoperability requirements.

The foundation for EDLM's understanding of DoD-wide data includes two fundamental items, a globally unique representation of a data term (the Tag) and the definition of the data term. The vocabulary (data terms) must support the tenets of Visible, Accessible, Understandable, Linkable, Trustworthy, Interoperable and Secure (VAULTIS). The development of a DoD-wide capability for managing DoD-wide identifiers and canonically controlled vocabularies is required to ensure consistency and clarity in the identification and utilization of resources across any legacy system connected to EDLM's data services. Multiple teams across the Office of the Secretary Defense (OSD) are currently working to enable prototype applications that deliver these capabilities. These prototypes are expected to set a

precedent for future enhancements and integrations within the EDLM program, paving the way for a more unified and efficient digital learning landscape.

DoDI 8320.03 addresses Unique Identification (UID) Standards for supporting the DoD Information Enterprise. It identifies several unique issuing authorities and places responsibility on functional leaders, without technical guidance to ensure a consistent or interoperable approach. These inconsistencies undercut efficacy in the enterprise and DoD's ability to consume data from any source with a certainty of identity. An enterprise of DoD's size must approach UIDs in a way that ensures data assets approach 100% certainty of their individual identity. UIDs are vital to all operational and analytical pursuits of the Department, including inventory (logistics), personnel (P&R), and decision support systems (advanced analytics including Artificial Intelligence disciplines), where each requires full confidence, the data is unique, discrete, and non-duplicative to enable accurate data-driven business and mission decisions.

DoD must also be able to manage our digital assets as they transit across the enterprise, having confidence in understanding their intent and meaning based on the origin (provenance) and each context (Domain) that has been verified as valid for use. The concept of a controlled vocabulary is necessary for almost all digital processes and functions and permeates the entire enterprise. When terms are not immediately and unerringly understood by every enterprise participant, there is a significant risk that process outcomes and analytic results will be inconsistent between communities. The lack of consistency and comprehension of each data term also significantly impacts the ability to aggregate, consolidate interoperate information in a meaningful way. The prototype-controlled vocabulary model ensures each EDLM data tag is machine-readable and is unique across the entire enterprise. This is necessary to assure a clear understanding of each specific term regardless of the originating mission or organization.

The implementation of unique identifiers (UIDs) and controlled vocabularies is crucial for the effective operation of data pipeline observers within the EDLM program. These elements serve as the backbone for accurately comparing real-time data generated at the point of delivery with the expected data outlined in the [x]BOM descriptions. The UIDs ensure each data element is distinctly recognized, while the controlled vocabularies provide a standardized language for describing the data, patterns, and canonical controlled vocabularies (CCVs). This harmonization is essential for building out data pipeline observers that can effectively monitor, analyze, and validate the integrity of data flowing through the EDLM systems, ensuring that the program's objectives are met with precision and consistency.

Conclusion

In terms of data impacts and outcomes, the [x]BOMs are anticipated to provide comprehensive insights into the interoperability of EDLM systems. The data collected will encompass various metrics, including system usage patterns, resource allocation efficiency, and the effectiveness of the training platforms in meeting educational objectives. This data will be instrumental in identifying areas for improvement, optimizing resource distribution, and ultimately enhancing the overall efficacy of the EDLM program. The integration of [x]BOMs within the EDLM program is not just a technological upgrade; it is a strategic move towards a more interconnected and harmonious digital learning ecosystem that supports the DoD's mission-critical objectives. As the program continues to evolve, the insights gained from [x]BOMs will undoubtedly contribute to shaping the future direction of enterprise digital learning within the defense sector.

ACKNOWLEDGEMENTS

The writing team for this paper would like to acknowledge the work being done by different participants in the DoD CIO working groups around [x]BOMS, UIDs, and Canonically Controlled Vocabularies. These working groups have provided technical insights into current best practices and operational requirements for creating, sharing, and maintaining [x]BOMs for all training and education applications in the future.

REFERENCES

1. **Department of Defense. (2016).** *DoD Instruction 1400.25, Volume 250, Civilian Strategic Human Capital Planning* [PDF file]. Retrieved from https://www.esd.whs.mil/portals/54/documents/dd/issuances/140025/140025_vol250.pdf.
2. Syafrizal, S., Selemat, & Zakaria. (2020, December). Analysis of cybersecurity standard and framework components.

3. European Commission. (2021, January 7). Industry 5.0: Towards more sustainable, resilient and human-centric industry. Retrieved from https://research-and-innovation.ec.europa.eu/news/all-research-and-innovation-news/industry-50-towards-more-sustainable-resilient-and-human-centric-industry-2021-01-07_en
4. SAP. (n.d.). Industry 5.0: Adding the human edge to industry 4.0. Retrieved June 11, 2024, from https://www.sap.com/insights/industry-5-0.html
5. Chief Digital and Artificial Intelligence Office. (2024). Data Mesh Reference Architecture (DMRA) (Version 2.6) [Draft]. Retrieved June 11, 2024, from https://media.defense.gov/2024/Mar/15/2003414274/-1/-1/1/dmra_paper.PDF
6. Haddud, A., DeSouza, A., Khare, A., & Lee, H. (2017). Examining potential benefits and challenges associated with the Internet of Things integration in supply chains. *Journal of Manufacturing Technology Management, 28*, 1055–1085. https://doi.org/10.1108/JMTM-09-2016-0132
7. Cybersecurity and Infrastructure Security Agency. (2023). A Hardware Bill of Materials Framework for Supply Chain Risk Management [PDF file]. Retrieved from https://www.cisa.gov/sites/default/files/2023-09/A%20Hardware%20Bill%20of%20Materials%20Framework%20for%20Supply%20Chain%20Risk%20Management%20%28508%29.pdf
8. Lass, S., & Gronau, N. (2020). A factory operating system for extending existing factories to Industry 4.0. *Computers in Industry*, *115*, 103128.
9. DaSilva, V.L., Kovaleski, J.L., & Pagani, R.N. (2018). Technology transfer in the supply chain oriented to industry 4.0: A literature review. *Technology Analysis & Strategic Management, 31*, 546–562.
10. Ahmed, K., Ramzan, Pevez, Azmat, Zeb, & Ur Rehman. (n.d.). Towards supply chain visibility using Internet of Things: A dyadic analysis review.
11. Khan, A., & Abonyi. (2022, December). Information sharing in supply chains – Interoperability in an era of circular economy.
12. Mageto, Prinsloo, & Luke. (n.d.). Determinants of logistics outsourcing performance among small and medium enterprises.
13. Stalnaker, Wintersgill, Chaparro, Di Penta, German, & Poshyvanyk. (2024, February). BOMs away! Inside the minds of stakeholders: A comprehensive study of bills of materials for software systems.
14. Belchior, R., Vasconcelos, A., Guerreiro, S., & Correia, M. (2021). A survey on blockchain interoperability: Past, present, and future trends. *ACM Computing Surveys, 54*(8), Article 168.
15. Hardjono, T., Lipton, A., & Pentland, A. (2019). Toward an interoperability architecture for blockchain autonomous systems. *IEEE Transactions on Engineering Management, 67*(4), 1298–1309.
16. U.S. Department of Education. (n.d.). The Family Educational Rights and Privacy Act (FERPA). Retrieved June 11, 2024, from https://www2.ed.gov/policy/gen/guid/fpco/ferpa/index.html
17. Vo, H.T., Wang, Z., Karunamoorthy, D., Wagner, J., Abebe, E., & Mohania, M. (2018). Internet of blockchains: Techniques and challenges ahead. In *2018 IEEE International Conference on Internet of Things (iThings), IEEE Green Computing and Communications (GreenCom), IEEE Cyber, Physical and Social Computing (CPSCom), and IEEE Smart Data (SmartData)* (pp. 1574–1581).
18. AVT Simulation. (2024). How to Build a Simulator Part IV: Elements. Retrieved June 11, 2024, from https://www.avtsim.com/how-to-build-a-simulator-part-iv-elements/
19. Dretzka, Fuller, & Smith. (2023). Weaving the [x]BOM fabric.
20. Azagury, Ashraf, Wright, Fang Grant, & Moore. (n.d.). Reinvention in the age of generative AI.

Fachada, D. (n.d.). Artificial Intelligence in Modeling and Simulation.