Enterprise Course Catalog

Pre-Alpha Prototype Final Report

31 August 2020

Distribution Statement A
Approved for public release: distribution unlimited.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
31-08-2020	Report	13-04-2020 - 31-08-2020

4. TITLE AND SUBTITLE

Enterprise Course Catalog
Pre-Alpha Prototype Final Report

5a. CONTRACT NUMBER

HQ003420C0109

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

0603769D8Z

6. AUTHOR(S)

Ashley Reardon, Brent Smith, Trey Hayden, Yihua Liu;
Systems Engineering and Technical Assistance (SETA) Contractors

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

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

Ashley Reardon, Brent Smith, Trey Hayden, Yihua Liu
Advanced Distributed Learning Initiative
13501 Ingenuity Drive, Suite 248
Orlando, Florida 32826

8. PERFORMING ORGANIZATION REPORT NUMBER

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

Assistant Secretary of Defense for Readiness
Advanced Distributed Learning Initiative
13501 Ingenuity Drive, Suite 248
Orlando, Florida 32826

10. SPONSOR/MONITOR'S ACRONYM(S)

OUSD/ASD-R/ADLI

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

Distribution A

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The 21st century is marked by significant technological progress in every field. For learning and development, these advancements have helped us realize the promise of “anytime, anywhere” learning as well as learning personalized to individual needs. More than that, emerging capabilities have thrown open the door to transformative possibilities, facilitating learning at scale, optimizing learning in response to large and diverse data sets, and developing fully integrated talent management systems for managing and enhancing the future workforce.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Nick Armendariz
U	U	U	UU		19b. TELEPHONE NUMBER *(Include area code)* 407-381-5550

Reset

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

Enterprise Course Catalog

Pre-Alpha Prototype Final Report

Prepared by
The ADL Initiative

DISTRIBUTION STATEMENT A. Approved for public release: distribution unlimited.

Table of Contents

Introduction **1**

Problem Statement 1

Solution 1

ECC Pre-Alpha Prototype Approach **3**

Work Breakdown Structure 3

Design Approach 5

Metadata Enumerations and Taxonomy 5

UI/UX Design 6

Development Approach 13

Next Steps **14**

Expanding on the Prototype 14

Aggregation and Persistence 14

Automation and Local Catalog Ownership 15

Coordination, Maturation, and Scale 15

Appendix **17**

Key Architectural Components/Definitions 17

INTRODUCTION

In July 2018, the DoD Chief Management Officer (CMO) and the Reform Management Group formally initiated the ***Enterprise Digital Learning Modernization*** (EDLM) reform initiative. The goal of the EDLM effort is to build an enterprise-wide integrated digital learning ecosystem that enables efficient acquisition and spending management for DoD education and training products and services. The ***Enterprise Course Catalog*** (ECC) is one of the three EDLM lines of effort supported by the Advanced Distributed Learning (ADL) Initiative.

Specifically, the ECC is a DoD-wide inventory of education and training assets that federate across local DoD networks and data sources, allowing personnel to use a single interface to access these resources. By providing a comprehensive catalog of online course content, the ECC will reduce duplications of effort for learners, training administrators, content developers, and acquisition personnel.

In July 2020, the ADL Initiative developed an ECC Pre-Alpha Prototype to validate the architectural approach for course catalog federation and to collect data about existing course catalog structures. Courses from Air Education and Training Command (AETC), Defense Acquisition University (DAU), and Naval Education and Training Command (NETC) were integrated into a common catalog portal powered by an open-source search and analytics engine (ElasticSearch) and visualization platform (Kibana).

This prototype effort revealed the need for a metadata standard to ensure harmonization across disparate data sources, and to enable the enrichment of metadata to improve the user experience for improved search and discoverability. This report summarizes the work to develop the ECC Prototype.

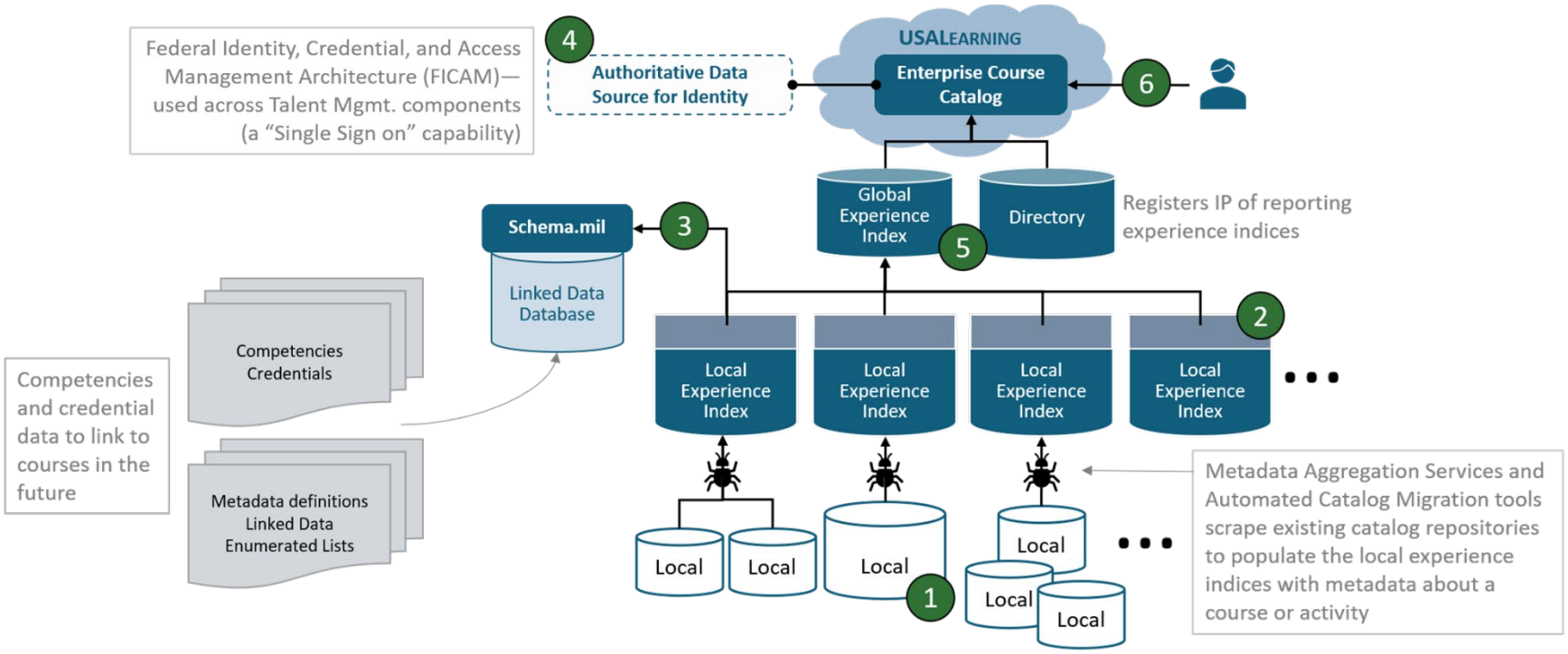
Problem Statement

To build a DoD-wide learning ecosystem, instructional activities and resources from across the agency must be discoverable and accessible. However, with hundreds of different education and training communities using a wide variety of methods to describe and publish their learning activities, there are now hundreds of proprietary and disconnected catalog capabilities across the DoD. The present lack of commonality and interoperability promotes inefficiency, duplication of efforts, and a lack of awareness about available learning opportunities.

Existing course catalogs are not architected to easily transfer data about learning activities between the different DoD systems that require these learner data. Catalogs integrated into proprietary platforms use pre-determined, point-to-point connections to transfer data between systems, which results in lengthy integration efforts for each connected system. These systems rarely accommodate new and emerging types of learning activities, such as e-books, mobile devices, augmented reality, or simulations, and they fail to provide insight into the learning activities that comprise each course. Current course catalogs also use sparse, non-standard metadata to describe their courses, which limits the ability to share resources across DoD components, and identify course duplication.

Solution

As seen in ***Figure 1***, the ECC enables a global search capability that pulls information from decentralized local catalogs across the DoD and aggregates the content into a single, Defense-wide catalog. In other words, existing course catalog systems and their respective contents will still be owned and managed by their current system owners, but the content (data) within those systems will be discoverable across all of DoD via a single web-based ECC portal.



(1) Air Education and Training Command (AETC), Defense Acquisition University (DAU), and Joint Knowledge Online (JKO) are collaborating on the development of the ECC by providing access to local catalog systems and supporting the ECC's testing and implementation. For the prototype, these existing catalogs are replicated in the Total Learning Architecture (TLA) developmental sandbox. The Initial Operational Capability (IOC) intends to provide the capability to begin migrating all course catalogs across the DoD. Full Operational Capability (FOC) will continue to polish ECC features and automated metadata generation capabilities. (2) Local course catalogs, content libraries, and other repositories will be indexed, and metadata will be automatically generated by the Metadata Aggregation Service and Automated Catalog Migration Tool. Those metadata will be stored within a local experience index. (3) Linked data will be stored in a schema server to provide shared vocabularies, such as for competency definitions and other standardized metadata attributes. (4) Federated Identity, Credential, and Access Management (FICAM) will facilitate a single sign-on capability by safely associating someone's personal identity with their system access permissions. (5) Local experience indices will roll-up to the global experience index. (6) End-users will be able to access content from this global experience index through a single, online web portal hosted by USALearning.

Figure 1 - ECC architectural approach

The ECC uses a metadata curation service to automatically generate metadata from different pools of information stored within the catalog owner's local network. For example, DAU requires students to complete a course survey at the end of each course. These data are stored in DAU's data warehouse and might be used to drive a course's aggregate rating, which is a common feature found in most commercial course catalogs.

The automated metadata service facilitates the alignment of metadata across organizations and helps maintain up-to-date information in the ECC. The metadata can also be used to align learning activities (e.g., courses, instructional materials) with standardized data on careers, competencies, and credentials. This kind of data alignment is a key pillar of the larger EDLM data strategy.

The ECC includes several subsystems that enable the discoverability and accessibility of all DoD instructional activities. The ADL Initiative is focused on building, integrating, testing, and deploying the services and systems required to enable this ECC vision.

ECC Pre-Alpha Prototype Approach

The ECC Pre-Alpha Prototype was developed to validate the existing architectural approach for course catalog federation and to collect data about existing course catalog structures. The prototype integrated existing data from three large DoD course catalogs, which set the baseline level of existing data maturity across DoD catalogs and demonstrated an ECC user interface (UI), called the *Digital Learning Portal* (DLP). ***Figure 2*** shows the prototype technical approach.

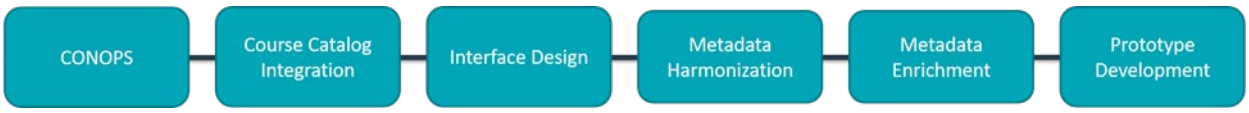


Figure 2 – ECC Prototype technical approach

The first step in our approach involved generating a Concept of Operations (CONOPS) to define characteristics of the ECC Prototype and establish the ECC scope. Next, three DoD course catalogs were obtained and integrated into an ADL Initiative sandbox environment. These raw data were referred to as 'Swamp' data and were represented in an open-source interface to demonstrate the current landscape of course catalog data components.

Next, use-cases were developed to demonstrate the expected user and system flow from representational users. Wireframes and mockups were then developed to design the UI and provide specifications to the development team. To further demonstrate the desired intent of the ECC, a small set of course catalog data were manually cleaned and enriched according to a common schema. This enabled the extraction of enriched representational metadata that aligned to the defined use-cases. Finally, the DLP interface was developed using an iterative agile software development process.

Work Breakdown Structure

A Work Breakdown Structure (WBS), depicted in ***Figure 3*** shows the major milestones, tasks, and sub-tasks that were required to develop the ECC Prototype.

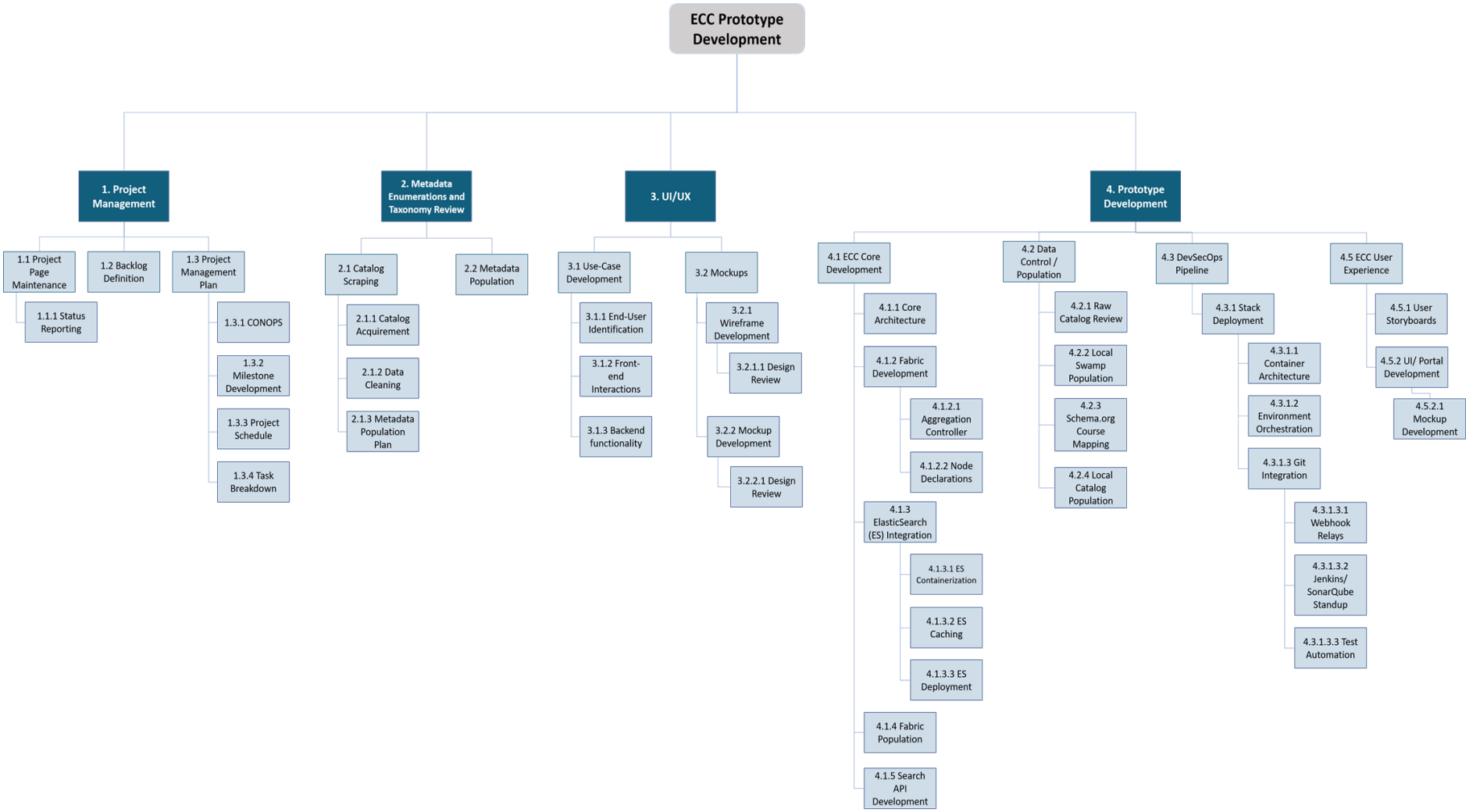


Figure 3 – ECC Prototype WBS

Design Approach

Given the relative complexity of prototyping an ECC, the team divided prototyping efforts into three categories: metadata processing, UI/UX design, and the core engineering work. Metadata efforts involved the ingestion of local catalog data and schema instrumentation of catalog entries. UI/UX efforts focused on DLP design, user stories, and coordination with engineering to ensure proper workflow for the target audience. Engineering efforts included all software architectures, cloud infrastructure, and systems integration.

Metadata Enumerations and Taxonomy

For the ECC prototype system, the ADL Initiative chose existing NETC, AETC, and DAU catalogs for the source material, based on factors such as the quantity, quality, and diversity of the data in each listing. These catalogs contained approximately 22,000, 3,500, and 27,000 entries, respectively.

These data sets were exported and delivered to the ADL Initiative in several different formats. The ADL Initiative staff uploaded the data to a central database in a common format. This required data extraction and cleaning from Adobe PDF files, Microsoft Access databases, and Microsoft Excel spreadsheets to create clean, interoperable CSV data sets.

After the raw data were extracted and cleaned, the ADL Initiative leveraged a common metadata schema (Schema.org) to enable interoperability and support expanded metadata fields. This was necessary as the catalogs exhibited significant differences. For example, the AETC catalog used several identifier fields per course, but lacked human-readable course names. One column was used for various details, such as registration information, prerequisites, and course descriptions, while another explained what type of information was recorded in the first column. This resulted in many courses with several entries in the catalog. ***Table 1*** shows the common-format approach used by the ADL Initiative.

Table 1 – Example schema application

Original Data Format			Schema
NETC	AETC	DAU	
Agency/Organization	Organization_Code	Agency/Org	maintainer
Course Provider	title	SOURCE/VENDOR	provider
Course_Title	[NOT INCLUDED]	Course Name	about
Course_Description	Paragraph_Text	COURSE DESCRIPTION	abstract
[NOT INCLUDED]	BCI_ID	Course ID #	courseCode
[NOT INCLUDED]	[NOT INCLUDED]	ITEM TYPE	additionalType
[NOT INCLUDED]	[NOT INCLUDED]	DELIVERY METHOD	encodingFormat

The NETC catalog data included agency name, course ID, name, and description, along with the format and delivery method. The AETC catalog was limited to electronic courses provided through a learning management system (LMS). DAU's iCatalog had similar fields but included several technical columns about things like Adobe Flash, Section 508 compliance, Experience API (xAPI), and SCORM. It listed intended audiences via abbreviations in a text column, the vendor and delivery method, and contact information for each course.

To aggregate and align the information from all three catalogs, the ADL Initiative mapped the fields from each catalog to Schema.org. Some fields were similar, but rarely did they represent a one-to-one mapping. Much of the information required to meet ECC's future metadata requirements was not available, though it existed outside the catalogs. For example, no catalog contained course ratings. As a result, some of the data were manually generated (relevant available information was used to create more realistic estimates wherever possible). ***Table 2*** shows how the metadata were formatted.

Table 2 – Example of manufactured data following Schema.org

Metadata Format	Metadata
additionalType	Professional Military Education Required Training
aggregatedRating	5
reviewCount	402
encodingFormat	In-Person
role	Trainee Student
cost	No cost
personnelType	Enlisted
component	Air Force
contentLocation	Arlington VA
completionTime	65
paygrade	E6 E7 E8

To illustrate, a critical use case involved the ECC behavior with respect to duplicate courses. The inconsistency across existing catalogs made it difficult to find duplicates. Even a standardized, widely used course, like the DoD Insider Threat Awareness training, was listed with different titles and descriptions. This makes finding an exact match significantly more complex and less reliable.

Often, a course is aligned with a *Terminal Learning Objective* (TLO), the desired end-result of the education activity. The ADL Initiative found minimal duplication at the TLO level, but it is likely that the migration to competency-based learning will reveal course duplication across *Enabling Learning Objectives* (ELOs) as they become integrated into competencies.

This demonstrates the importance of a common metadata standard that can be applied throughout DoD. While there are many existing e-learning metadata standards, they often fail to address the wide range of use cases a modern learning ecosystem serves. Through efforts of the ADL Initiative's Total Learning Architecture (TLA) working group between 2018 and 2020, the ADL Initiative collaborated with the Institute of Electrical and Electronics Engineers (IEEE) and other stakeholders across industry, academia, and government to begin defining metadata requirements to support artificial intelligence, machine learning, and big data analytics required to support future learning technologies.

This work resulted in a new IEEE P2881 Learning Activity Metadata study group that is working toward standardizing an approach that maintains interoperability with legacy standards, including Learning Object Metadata (LOM), the Learning Resource Metadata Initiative (LRMI), and Schema.org, while offering greater support for newer learning modalities, such as simulators and virtual reality.

UI/UX Design

The DLP interface design was created from the CONOPS and use-cases developed based on interviews with stakeholders identified by the ADL Initiative. It reflects a typical workflow from the perspectives of both learners and instructors. The DLP, set up by the ADL Initiative, is comprised of a landing page, advanced search, search results, and detailed course views. While future iterations may include greater functionality, an ECC 90-Day Sprint focused on building out the following three use cases.

Use-Case #1	Searching for Annual Training Courses	
Primary Actor	Learner	
Description	The Learner is required to take the 'Ethics at the Operational Level' course in order to fulfil their annual training requirements.	
Preconditions	The Learner has a registered account with the DLP and is signed in.	
Trigger	The use-case is initiated when the Learner accesses the DLP.	
Typical Course of Events	**Actor Action**	**System Response**
	Step 1: Learner navigates to the DLP.	Step 1.1: Search bar is visible. 'Popular' dashboard is displayed.
	Step 2: Learner uses the scroll bar to search through the 'Popular' dashboard.	Step 2.1: Select courses are previewed according to their current popularity. Course bounding box is clickable.
	Step 3: Learner selects the 'Ethics at the Operational Level' button.	Step 3.1: User is redirected to the detailed course description page for 'Ethics at the Operational Level'.
	Step 4: Learner selects 'Enroll'.	Step 4.1: Redirects to Moodle where registration can occur. * **Current functionality not developed.*
Conclusion	The use case concludes when the learner selects the 'Enroll' button in the 'Ethics at the Operational Level' detailed course view.	

Use-Case #2	Searching for Voluntary Learning Opportunities	
Primary Actor	Learner	
Description	The Learner is looking to improve their acquisition skills. Specifically, they are looking for an in-person acquisition course that is available in their current town.	
Preconditions	The Learner has a registered account with the DLP and is signed in.	
Trigger	The use-case is initiated when the Learner accesses the DLP.	
Typical Course of Events	**Actor Action**	**System Response**
	Step 1: Learner navigates to the DLP.	Step 1.1: Search bar is visible. 'Popular' dashboard is displayed.
	Step 2: Learner selects the 'Advanced Search' link below the search bar.	Step 2.1: System redirects to the 'Advanced Search' page, presenting several options for filtering the users search.
	Step 3: Learner types in 'Acquisition' in the keyword search box. // Learner selects 'Open' in the 'Availability' box. // Learner selects 'Portsmouth, VA' in the 'Location' box. // Learner selects the 'Advanced Search' button.	Step 3.1: System navigates to the search results page. Results are prioritized according to filtering selections.
	Step 4: Learner browses options and selects 'SPAWAR Risk Management Framework (RMF) Risk Assessor Course: SCA – 201)'.	Step 4.1: User is redirected to the detailed course description page for 'SPAWAR Risk Management Framework (RMF) Risk Assessor Course: SCA – 201)'.
	Step 5: Learner selects 'Enroll'.	Step 5.1: Redirects to Moodle where registration can occur. * *Current functionality not developed.
Conclusion	The use case concludes when the Learner selects the 'Enroll' button in the 'SPAWAR Risk Management Framework (RMF) Risk Assessor Course: SCA – 201)' detailed course view.	

Use-Case #3	Identification of Duplicates	
Primary Actor	Instructor	
Description	The Instructor just received notification that their students must receive annual training on insider threat awareness. They are looking for similar courses that they can reuse.	
Preconditions	The Instructor has a registered account with the DLP and is signed in.	
Trigger	The use-case is initiated when the Instructor accesses the DLP.	
Typical Course of Events	**Actor Action**	**System Response**
	Step 1: Instructor navigates to the DLP.	Step 1.1: Search bar is visible. 'Popular' dashboard is displayed.
	Step 2: Instructor types 'Insider Threat Awareness' in the search bar.	Step 2.1: System navigates to the search results page. Results are prioritized according to keyword fit.
	Step 3: Instructor browses options and selects 'Insider Threat Awareness'.	Step 3.1: User is redirected to the detailed course description page for 'Insider Threat Awareness'.
	Step 4: Instructor reviews course information.	N/A
Conclusion	The use case concludes when the instructor is redirected to the 'Insider Threat Awareness Course' and assesses the course for reuse.	

Landing Page

As shown in ***Figure 4*** on the following page, current access to the DLP is accomplished by logging in with a TLA portal log-in. Future capabilities will provide role-based logins from the perspective of learners, non-learners, and administrators. Metadata from the courses, such as category, aggregated ratings, and modality, might be used to drive the 'Popular' feature. A 'Resources' tab might include links to other relevant resources, such as job aids, competency and credential management resources, or ancillary learning content such as eBooks, mobile learning applications, serious games, or simulation scenarios and exercises. The search bar is prominently placed, immediately allowing users to engage with the course catalog.

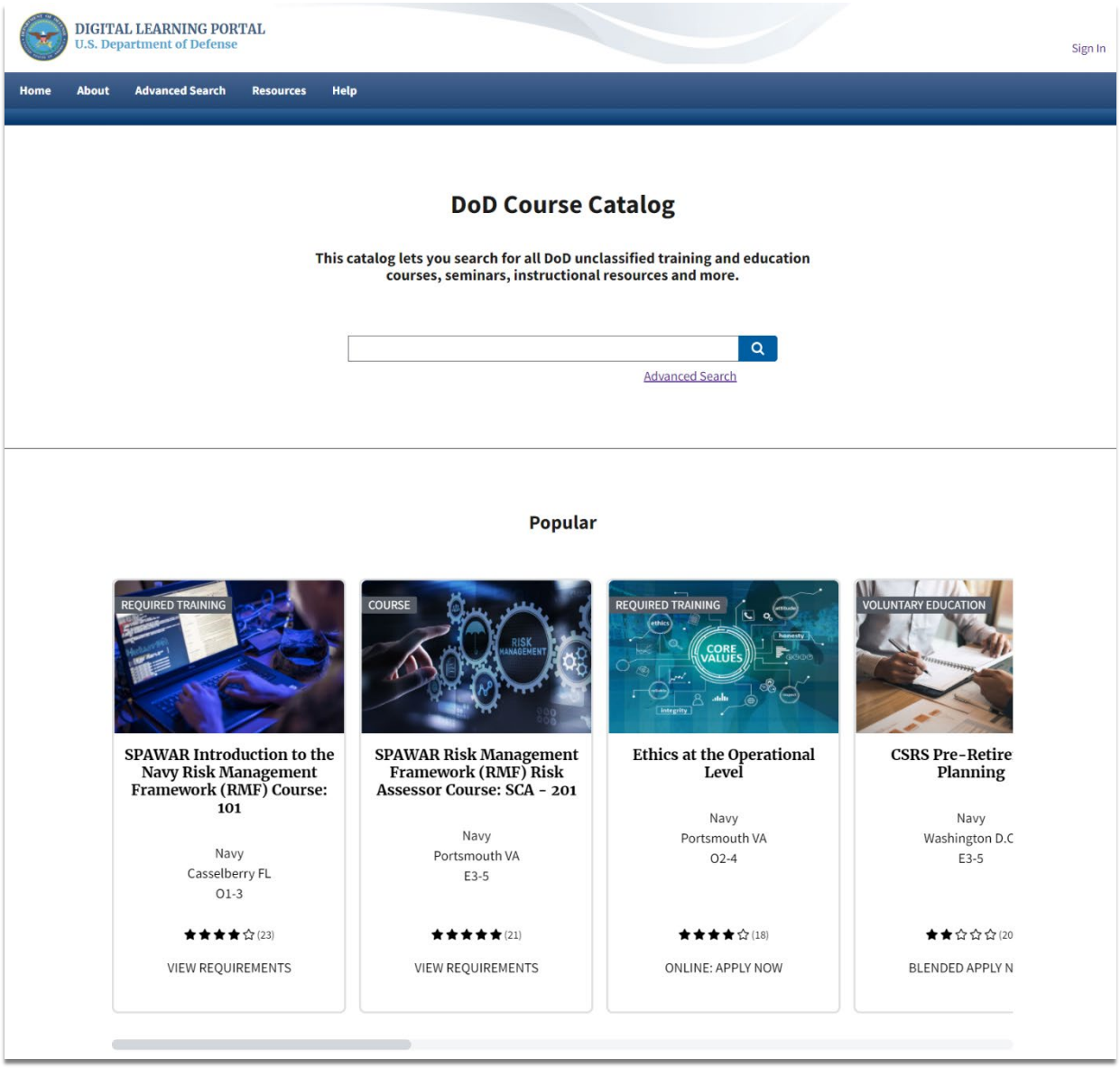


Figure 4 – ECC Digital Learning Portal

Advanced Search

The Advanced Search feature, shown in ***Figure 5***, provides several ways to narrow down course searches based on enriched metadata. By using a common metadata standard, representational data was included to search upon those fields. As discovered in early metadata analysis, these metadata are not typically included in course catalogs, but might exist elsewhere in the DoD such as in training performance requirements documents. Inclusion of these metadata can enable a more customized approach to course search and discoverability, which may ultimately reduce time looking for education and training opportunities.

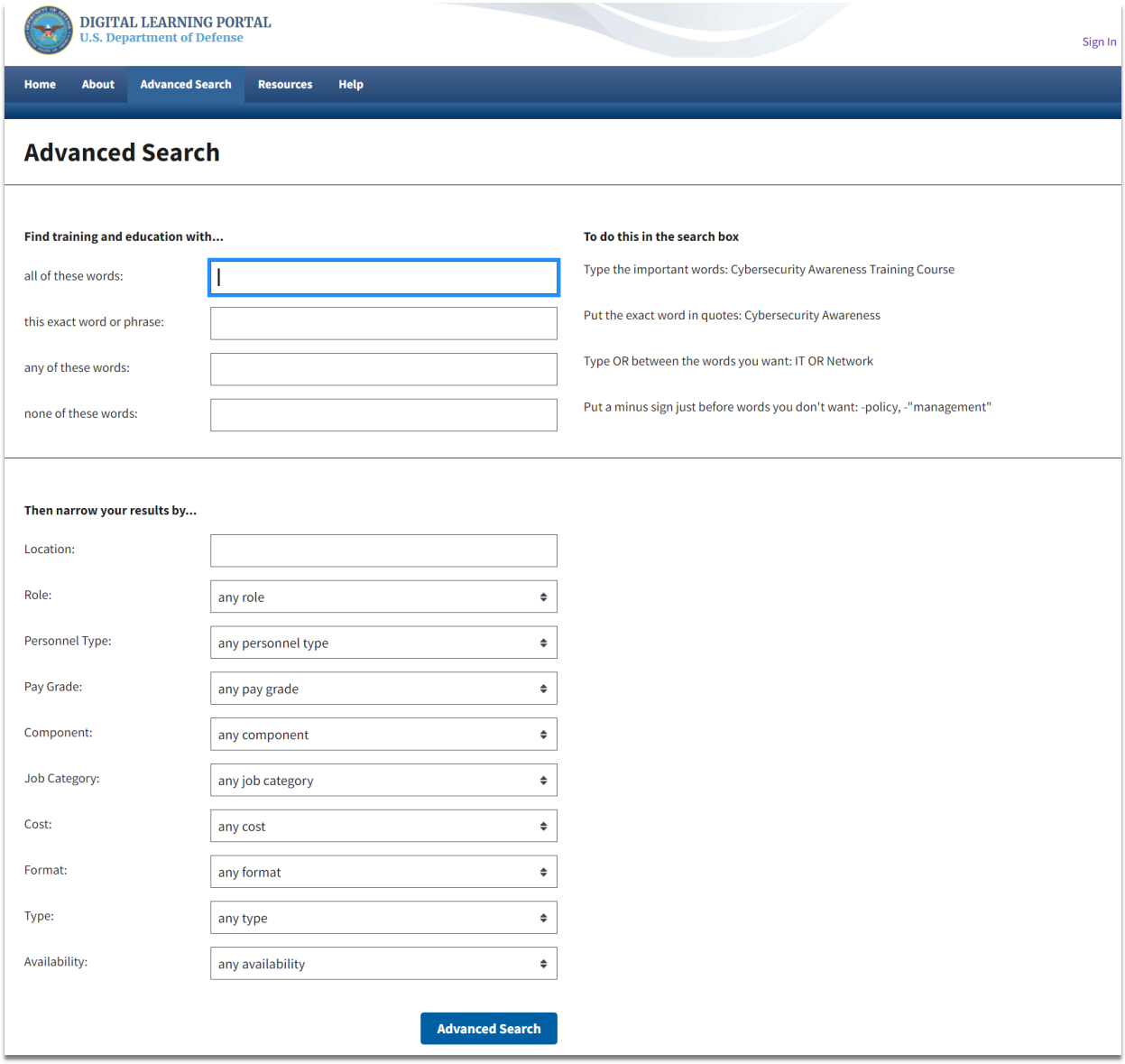


Figure 5 – ECC Advanced Search

Search Results

The Search Results page, shown in ***Figure 6***, highlights the application of these metadata in providing information of relevance for narrowing down searches. A filter bar was included in this page to afford users the opportunity to further narrow their search without having to navigate back to the Advanced Search page.

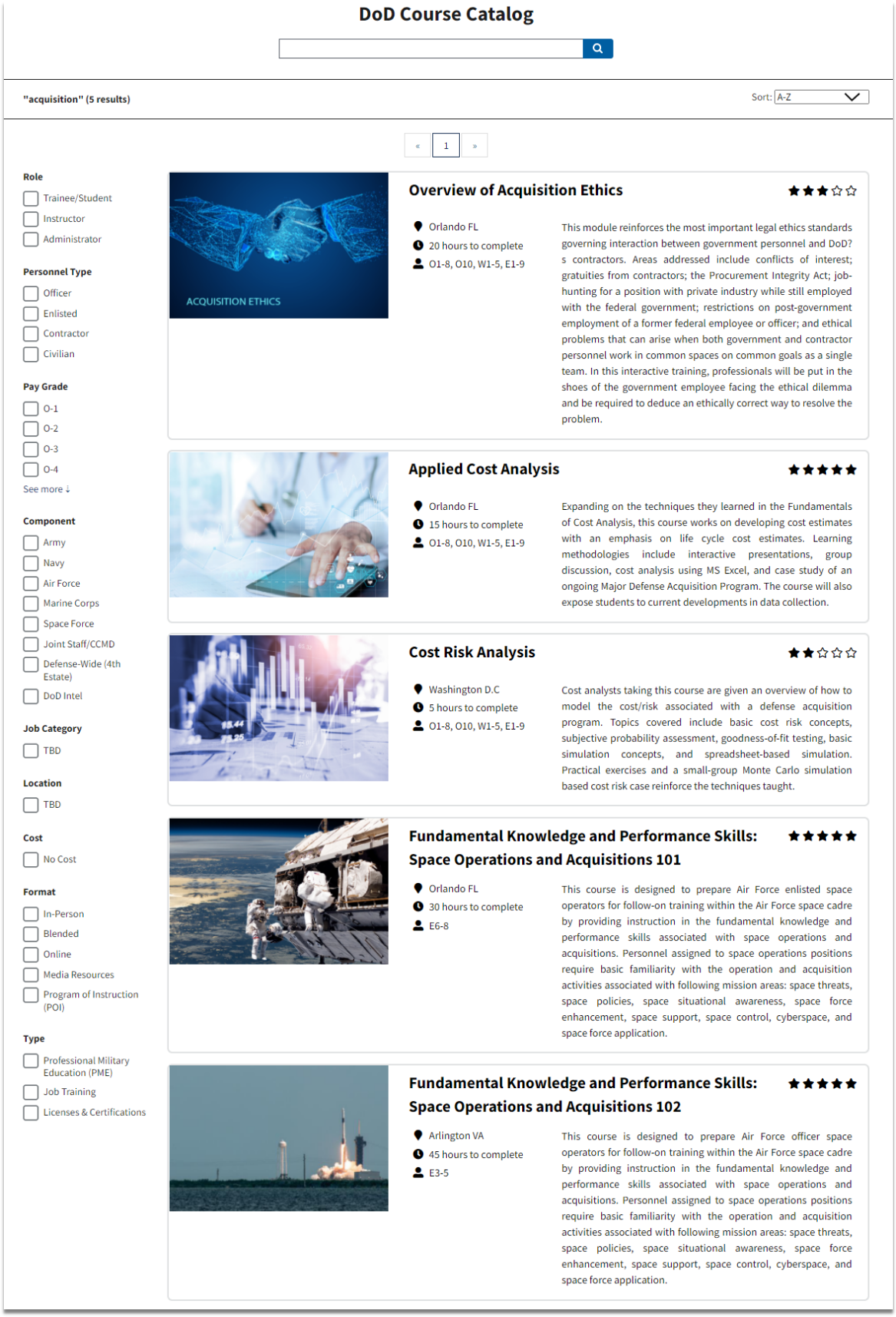


Figure 6 – ECC Search Results

Detailed Course View

The detailed course description page, shown in ***Figure 7***, provides a method for communicating pertinent course information for a variety of end-users. Future capabilities might include course evaluation reports when logged in as an administrator or course manager and might include features for easily assessing duplicate courses across the enterprise.

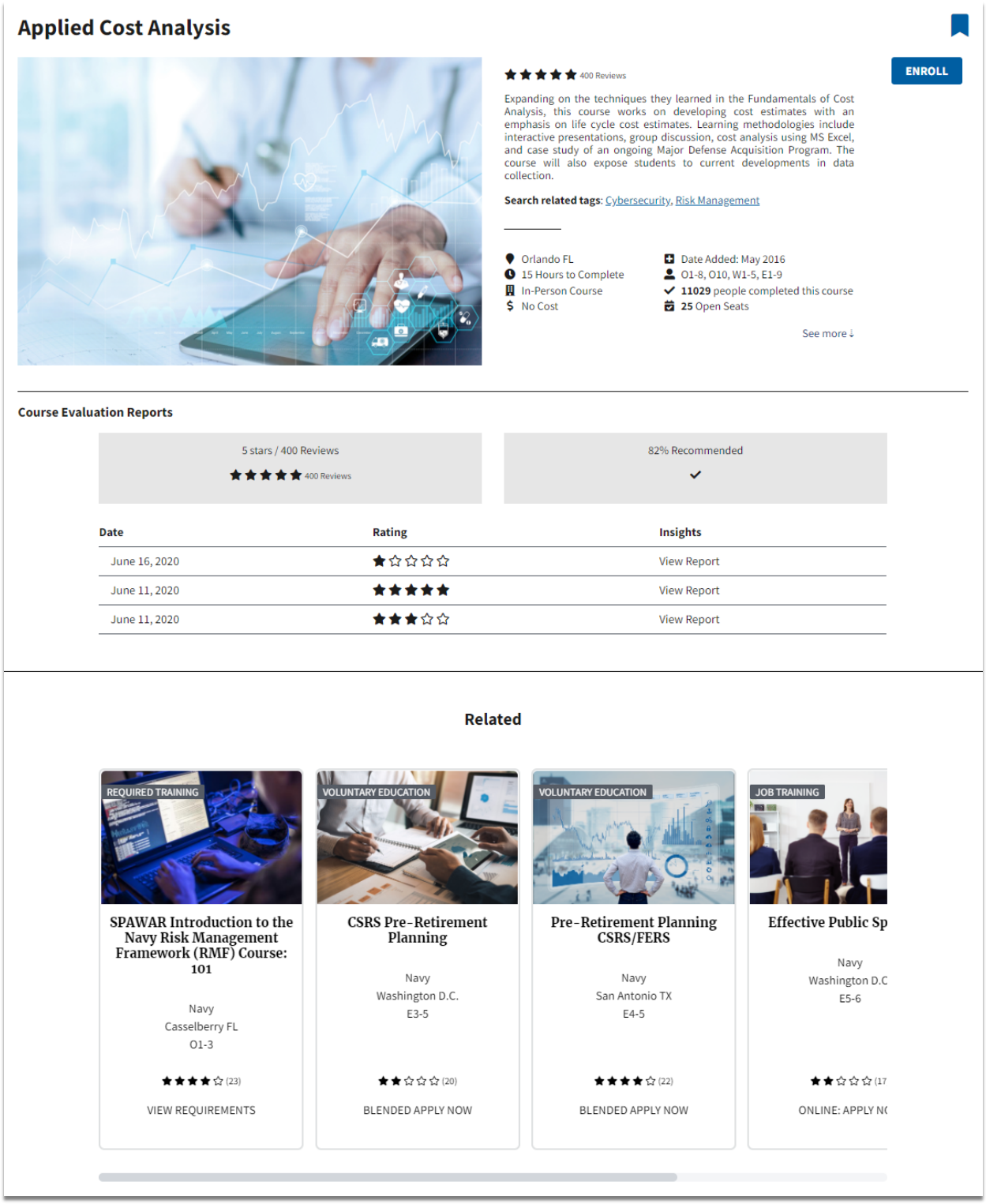


Figure 7 – ECC Detailed Course View

Development Approach

ADL Initiative engineers developed the ECC Prototype using an agile approach that divided ECC development into phases called *Swamp* and *Mockup*. Both phases required complete architectures with a set of local catalogs, a data fabric, and a portal, which allowed rapid testing and deployment of individual components. The Swamp phase was straightforward and included the instantiation of ElasticSearch and Kibana as critical foundations of the prototype. The Mockup phase included refining the Swamp data and producing a newer DoD-themed portal to demonstrate how a mature ECC capability might behave.

Local Catalogs

For the prototype, the team received local catalog data sets from NETC, AETC, and DAU. Aside from their differing file formats, these catalogs used conflicting data schemas, which required the use of document-style data stores rather than a traditional table format. The team used MongoDB as a database service, stood up three instances with simple APIs for reading their catalog entries, and wrote a simple comma-separated values (CSV) upload script to populate these instances with the serialized catalogs. After configuring access credentials, the team configured the data fabric to target and aggregate these catalog replicas.

As the catalog effort transitioned between the Swamp system and the ECC DLP mockup, a standardized metadata format was required. A small portion of the original catalog entries were modified to use Schema.org's course structure in addition to their unique fields. Once these fields were present, the team uploaded the updated entries to the existing MongoDB instances and configured the fabric to dynamically choose between Swamp and the cleaned Mockup data.

Data Fabric

The ECC architecture uses a data fabric approach for federating the local course catalogs. Generally speaking, data fabrics are strategies and systems for aggregating, harmonizing, and serving otherwise disparate data. This concept appeals to the ECC's most obvious objective of federating siloed catalogs and presenting them through a single user interface. For this, the team developed a simple fabric architecture and broke its responsibilities out across two subsystems: *aggregation* and *search*. Future iterations should rely on the IEEE P2881 Learning Activity Metadata, which will build upon Schema.org to meet DoD specific requirements.

Aggregation was simplified considerably by the use case. As the catalogs did not require real-time, instantaneous federation from source to user, local catalogs could be pulled and indexed within ElasticSearch without needing to manage search indices and sessions across those catalogs during runtime. Those catalogs were instead pulled by an aggregation service (named Grendel) and indexed directly to the ElasticSearch node. As that node was part of a containerized ElasticSearch/Logstash/Kibana (ELK) stack, the aggregated catalog entries could be viewed in-browser through the Kibana interface, allowing the team to confirm the process functionality.

The aggregation workflow was functional but not perfect. Specifically, modification of records would naturally be done in their source catalogs, but this necessitated a re-indexing of all modified catalog entries. The team set this issue aside during the 90-day effort due to time constraints. A more robust editing and re-indexing strategy will be critical in future ECC implementations to preserve an entry's original URL and record ID.

The search function was largely handled by ElasticSearch. While ElasticSearch performed the actual search logic and had excellent documentation for its API, some of the ECC filtering requirements (e.g. pay grade) had verbose inbound formats requiring pre-processing prior to sending the ElasticSearch query. The search result payload from ElasticSearch already contained metadata regarding the search process and its results, so the team modified that payload to assist with UI pagination and trimmed unnecessary information before returning the Search API's response.

Mockup Portal Development

The effort included development of a user interface for searching ECC entries. As the team already had an ELK stack for ElasticSearch, the open-source tool Kibana served as a premade administrator UI and enabled the team to confirm the aggregation system's functionality. In parallel, the team developed a more formal end-user UI – a Jekyll site following Federalist guidelines, using the US Web Design System (USWDS) style guides and traditional DoD branding. After populating the Jekyll mockup site with dummy data, the team fielded and integrated feedback from test users to identify issues, confirm accessibility across devices, and ensure an overall positive user experience.

Once the mockup design passed initial user testing, the team began integrating the UI with the actual search data. After migrating the mockups from Jekyll into a more familiar framework, the portal server could retrieve catalog entries from the Search API and present those entries to the user in-browser, (this also served as a means of iteratively testing the Advanced Search functionality). The team limited and removed some features (e.g. Administrative View) during development due to project resource constraints. When development ended, the portal could navigate ECC entries, support advanced search filtering, and demonstrate a baseline feasibility to the team's overall ECC approach.

NEXT STEPS

Before moving from the prototype architecture and implementation to a more robust solution to serve the larger DoD community, the team identified issues with the current approach that warrant more design work:

Expanding on the Prototype

While the prototype implementation yielded a functional, distributed catalog system, the team cautions against building upon this implementation directly. Rather, future teams may benefit from simply understanding the architectural approach and the implementation tradeoffs from this system. This prototype effort serves as a learning exercise for developing a more robust system that can account for its shortcomings and provide better value for DoD.

Aggregation and Persistence

As the catalog entries within the ECC must be replicas of the local catalogs' contents, an ECC solution will need to ensure both the accuracy of its aggregated catalog and the persistence of any links to catalog entries within the ECC portal. The prototype did not handle this well, as each ElasticSearch index was simply cleared and replaced with each aggregation, and its portal links were based on the internal ElasticSearch document index. A robust system could use the pairwise uniqueness of the catalog source and the catalog entry's fields when providing ECC links, or it may simply leverage an existing internal ID unique to the local catalog store.

Automation and Local Catalog Ownership

While the aggregation from local catalogs seems like an inevitable architectural decision, ownership of the local catalogs must be maintained by current catalog owners. This decision impacts not only the general connection authority requirements for such a network but also the necessary sophistication of automated scraping tools to assist with that aggregation – particularly important when considering the amount of data being aggregated.

Coordination, Maturation, and Scale

Given the eventual ECC scope (federating catalogs across the entire DoD), ensuring that this architecture and its subsequent implementation(s) can succeed at-scale is paramount. This effort goes beyond what the ADL Initiative can accomplish by itself and its success will require coordination among the EDLM partners and other stakeholders in Government, industry, and academia where the larger development and implementation work will rest. Schema standardization, aggregation authorities, performance at-scale, and feasible policy are all required for this effort to succeed.

Table 3 provides a Capability Maturity Model (CMM) to gauge the anticipated levels of achievement and functionality for the ECC project. The project currently is approaching CMM Level 1.

Table 3 - ECC Technical Capability Maturity Model

Level	0 (Baseline)	1 Prototype	2 IOC	3 (Integrated) FOC	4 (Federated) TLA
Semantic Interoperability	None, walled garden systems with no protocol or semantic interoperability.	Formal course content is constrained. Not all fields are filled out. Ancillary (e.g., annual training) and curated content (e.g., instructor manuals) can't be easily accessed.	Formal content adheres to DoD metadata standards. Ancillary (e.g., annual training) and curated content (e.g., instructor manuals) can't be easily accessed.	Ancillary (e.g., annual training) and curated content (e.g., instructor manuals) can be accessed. No common standard for semantic interoperability.	Ancillary (e.g., annual training) and curated content (e.g., instructor manuals) can be accessed. Adheres to common standard for semantic interoperability.
Metadata Strategy	Formal course content is constrained to documents (e.g., Data Item Description (DID), PDFs, Word Document), and is not currently being leveraged.	Job Duty Task Analysis (JDTA) or equivalent (e.g., job duty task analysis, maintenance task analysis, work task analysis, job task analysis, etc.) data is imported.	Initial integration of a formal DoD metadata strategy driven by DIDs. Expands to include Ancillary (e.g., annual training) and curated content (e.g., instructor manuals). Metadata scraping of the DIDs occurs to extract content.	Ancillary (e.g., annual training) and curated content (e.g., instructor manuals) clearing house is built to enable access of content beyond the course level. Automatic metadata scraper and services automatically imports all formal course data and ancillary and curated content to access supporting course information and information beyond the course level.	Metadata is automatically linked to competencies and credential data through a Schema.mil server.
Metadata Extraction	Information about formal course content is populated manually.	Formal course metadata is manually transported from highly-typed (pro forma) documents, like DIDs. (e.g., Instructional Performance Requirements Document (IPRD)).	Ancillary (e.g., annual training) and curated content (e.g., instructor manuals) metadata is manually entered.	Metadata Aggregation Services and Automated Catalog Migration tool scrape ancillary and curated content. Metadata is automatically populated.	Linked data is stored in a schema server to provide shared vocabularies for competency definitions and other standardized metadata attributes. A Linked Data Database provides a repository of linked data, namespaces, and vocabularies to add context to different sets of learner data.
Search and Discoverability	Search and discoverability is limited to legacy catalogs within an organization. Information is not correlated to your starting point; therefore you must know exactly what you are looking for prior to your search. There is no formal way to share ancillary (e.g., annual training) or curated content (e.g., instructor manuals).	Transfers for reading data about formal courses are established. Local Experience Indices are established.	Transfers for reading data about ancillary (e.g., annual training) and curated content (e.g., instructor manuals) are established. Local Experience Indices are established. Registration and discovery services are added.	Cybersecurity metadata scraping components are integrated. Federation to the Global Experience Index is enabled (e.g., NIPR > W3).	Integration with TLA components (e.g., ELRR) enables role-based login to a common portal that houses Training and Education content across the DoD. Content is customized according to your starting point (e.g., career state).

APPENDIX

Key Architectural Components/Definitions

Portal

The ECC portal provides role-based log-in for learners, instructors, and education/training administrators. It provides a common access point for accessing information about all courses, classes, and supporting activities across the DoD. For initial testing, the portal may use dummy logins. The Initial Operational Capability shall include appropriately secure and managed logins, and the Full Operational Capability shall include DoD Common Access Card (CAC) and single sign-on access.

DoD Schema Server

The schema server provides an enterprise linked data capability for describing ECC resources and aligning these resources to common vocabularies and any terms used within the definitions of competencies and credentials. While competency frameworks may not be available for all courses during the development effort, the ADL Initiative will provide representational data to satisfy the intent of the demonstration.

Experience Index

An Experience Index stores metadata and other information about an organization's local course catalog and aggregates it into a federated data repository that can be accessed via a metadata aggregation service.

Metadata Aggregation Services

The ECC leverages the existing metadata aggregation service used in the Open edX platform for storing and caching metadata about a course. This results in an ECC-specific implementation that uses a common metadata strategy provided by the ADL Initiative (derived from the draft P2881 Learning Activity Metadata standard being submitted through IEEE).

Automated Catalog Migration Tool

A data migration tool shall scrape the existing catalog repositories (i.e., provided by the NETC, AETC, and DAU stakeholders for the prototyping project) to populate the local Experience Indices. This capability will demonstrate limited semantic abilities to populate metadata fields with existing catalog data and dissimilar naming conventions. Not all metadata is expected to be populated for the prototype, but it should be designed to support metadata for when it is.

Metadata

Data that provides information about other data. The ECC uses metadata as a foundation for categorizing, enabling search functions, and accessing qualitative and quantitative information about learning activities, resources, and learner performance.