

Tell Me About It: Saving the Context of Archived Military Imagery

Thomas M. Ruyle & Julia Hickey, Defense Imagery Management Operations Center (DIMOC), Defense Media Activity (DMA), Ft. Meade, MD

Abstract

Defense Media Activity (DMA) is the Department of Defense's (DoD) direct line of communication for news and information to U.S. forces worldwide. The agency informs DoD audiences, entertains DoD audience overseas, trains Public Affairs and Visual Information professionals, and manages the DoD's visual information. Defense Imagery Management Operations Center (DIMOC) provides DoD enterprise-level visual information services including operational support, digitization, storage, access, records management of the Department's visual content, and accessions to the U.S. National Archives and Records Administration.

Since the mid-1990s, DIMOC has sought to digitally catalog, archive and make public millions of relevant still and video images taken by Department of Defense (DoD) photographers around the world. The images depict everything from mundane 'day-in-the-life' events to combat operations; many of the cataloged images will find their way into National Archives collections.

In addition to sharing unclassified images with the American public, DIMOC is also tasked with ensuring current operational imagery is available to decision-makers at all levels of the DoD.

The purpose of this paper is to describe in detail the improvements to our processes as well as the challenges that remain, and how DIMOC has evolved to meet the challenges of preserving our nation's military pictorial history.

Background Information

Defense Media Activity (DMA) is the Department of Defense's (DoD) direct line of communication for news and information to U.S. forces worldwide. The agency informs DoD audiences, entertains DoD audiences overseas, trains Public Affairs and Visual Information professionals, and manages the DoD's visual information. DMA informs millions of active, Guard and Reserve service members, civilian employees, contractors, military retirees and their families in the U.S. and abroad. Defense Visual Information (DVI) manages DoD visual information in support of U.S. military activities and operations and conducts visual information planning, policy, procedures, guidance, management, and standards. Defense Imagery Management Operations Center (DIMOC) provides DoD enterprise-level visual information services including operational support, digitization, storage, access, records management of the Department's visual content, and accessions to the U.S. National Archives and Records Administration.

Historical summary

Since the earliest days of military photography, perhaps the biggest frustration for those who use such photos for research is a lack of supporting information about the photo itself. Far too often, DIMOC encounters current and historical imagery without basic background information, such as a date, location or even a short caption identifying the subject of the photo.

Such imagery, while interesting and often pleasing to look at, holds little value to historians, the media or anyone else who would be interested.

In order for the images to be universally relevant, the primary responsibility for accurate, informative, context-driven captions and supporting metadata falls on the person who took the images. As the process of ingesting, cataloging and curating the imagery moved into the digital age, DIMOC established numerous metadata fields to more accurately search for assets from a particular command, location, date range, military service, operation or piece of equipment, among many other variables.

However, even with the technological progress of the past 30 years, it became clear that photographers still submitted many images with scarce metadata and sometimes even less context within captions.

To that end, during the late 1990s, DIMOC began developing what eventually became the Department of Defense Captioning Style Guide (DoDCSG) [1], which documents the importance of well-written captions, metadata and context, and provides guidance for successful submission of imagery to DIMOC for processing and preservation.



Figure 1. An example of low-value caption and metadata. This image only has the caption, "VIP-32, Jimmy Stewart." [2] No date, location or contextual information is available.

Getting it right

The first editions of the style guide were short PowerPoint presentations, covering the very basics of captioning (Who, What, Where, Why, When and How). To expand the guide's scope and effectiveness, DIMOC worked with instructors at the Defense Information School (DINFOS), where military journalists, photographers and public affairs personnel are trained in the production of military visual information.

The DoDCSG was officially created in 2010 [3] and DIMOC hired a Lead Writer-Editor, charged with developing a formal written style for captions and metadata to be applied to military photography across the Department of Defense. The first sanctioned edition of the

DoDCSG was released in November 2011 and has been updated several times. The DIMOC Joint Combat Camera Center (JCCC), which plans and monitors the flow of imagery from the field and fleet, distributes the DoDCSG to military photographers around the world.

In order to ensure captions are understandable by military and civilian audiences alike, DIMOC uses the tenets of the Associated Press (AP) Stylebook [4] – the news-writing style employed in most U.S. news outlets – as the basis for punctuation, grammar and spelling. The DoDCSG specifies proper usage of many military terms not covered by AP, and provides guidance on the various metadata fields used by the DoD.

The DoDCSG makes the following key instructional points with regard to good caption writing:

- Don't use military jargon or acronyms – members of the public, and even members of other military services, may not know what a particular acronym means.
- Spell out military ranks – for example, not many people outside of the DoD would know that a Navy PO1 is a Petty Officer 1st Class.
- Correctly identify weapons, aircraft, military units, ships (with hull number) and other equipment.
- Ensure dates and locations are correctly noted in the caption as well as within other metadata points.
- Identify all principal subjects within an image, as well as any action those persons are performing.
- Perhaps most importantly, context: Provide a basic background as to why the action in the image is happening. For example, the subject(s) of the photo were participating in a specific exercise or operation.

Furthermore, the DoDCSG outlines the metadata fields and specific information required in order to effect proper recordkeeping of the millions of images in DIMOC's holdings.

Perhaps the two most important fields, outside of the caption itself, are the Visual Information Record Identification Number (VIRIN), and keywords used for search engine optimization (SEO).

The VIRIN traces its history to the mid-1990s, when it became apparent that each image needed a unique, easily searchable identifier within the rapidly expanding digital archives. The VIRIN was initially designed as a four-field data string, with the following fields separated by hyphens:

- Field 1, six characters: The date the image was created, in YYMMDD format.
- Field 2, one character: A code corresponding to the originating military service or other organization.
- Field 3, five characters: A unique identifier for each military photographer. This was originally the initial of the shooter's last name, followed by the last four numbers of his/her Social Security number. In 2010, given concerns over personal information, this was changed to a two-letter, three-digit identifier, called the Vision ID, assigned specifically and permanently to each shooter through DIMOC.
- Field 4, three characters: A frame number to identify specific shots taken on a given day.

For example, the VIRIN for the ninth submitted photo generated by a specific Army photographer on June 25, 2013, would read 130625-A-LU685-009.

Keywording has become central to increased searchability of DIMOC-held assets; this is where a photographer can enter data for

unit names and abbreviations, service names, military occupational specialties, aircraft nomenclature, ship types and other specific items and themes depicted in an image. Keywording is undergoing a sea change, and it will be addressed later in this document.

Other metadata fields covered in the DoDCSG include: Operation/Exercise name (as applicable); Location fields (base/city, state/province, country/body of water); Release instructions (imagery released to the public must have proper vetting and the name, unit and contact information of the authorized release authority); Photographer's name and rank; Unit supported; Command Shown (one of the six unified Combatant Commands); and Service Shown (Not the photographer's military affiliation, but the service depicted in the image itself – military photographers often document the activities of other services).

Curation efforts

In 2010, DIMOC initiated a five-year contract to have a team of professional editors inspect, review and revise captions and metadata of all new imagery entering the Defense Asset Management System (DAMS). While the quality of the end product – approximately 600,000 individual archived still and video images – improved dramatically, it became clear this system of 'surgical' editing was not cost-effective.

Over the course of the contract, however, with the advent of the DoDCSG and a lot of coaching of DoD photographers by the JCCC, the overall quality of raw captions and metadata from the field improved greatly overall. The Lead Writer-Editor continues to field questions regularly and conduct classes for deploying photographers, in addition to working with DINFOS to stress the importance of captions and metadata to photographer trainees.

Workflow

The JCCC coordinates with military photographers at all levels within the various military structures, ensuring that current, high-demand imagery is available to DoD decision-makers, other government agencies, the media and the general public. Once a planner at the JCCC is aware of an upcoming event, he or she will ensure the photographers documenting the event have the right tools to submit the imagery in a timely fashion; the planners also ensure captions and metadata are written in accordance with the DoDCSG.

DIMOC personnel select imagery for the customers mentioned above on a daily basis as well as respond to specific requests for imagery.

Examples of specific requests would be for all archived images of a former President upon his death, or all images from a newsworthy issue, such as Hurricane Sandy in 2012. JCCC planners and military personnel take an active role in reviewing high-demand images for proper captions and metadata.

Most imagery is submitted to DIMOC via the Defense Video and Imagery Distribution System (DVIDS). DVIDS is a digital clearinghouse for unclassified, released DoD media (stills, video, news stories, audio, etc.) to be made available to the public and the media.

Shortly after ingest at DVIDS, the still and video imagery is further transmitted to a content management system (CMS) contracted by DIMOC for permanent storage. Since October 2014, DIMOC has curated imagery within this contracted CMS; DIMOC is slated to start working within DVIDS to curate metadata starting in early 2016, with the goal of sending only properly curated imagery to the contracted CMS.

The military services over the past several years have recognized the need for well-curated archival imagery, and are

beginning to take a greater share of the responsibility for ensuring the standards are met before images are transmitted to DVIDS. DIMOC has been working with the services on standardizing the curation process DoD-wide, but as with any government operation, change can be daunting and slow.

Special problems

While DIMOC has noted a vast improvement in the overall quality of captions and metadata submitted from the field, there are always circumstances in which the standards are not met.

One of those problems is when photographers – often an untrained person temporarily detailed to document an event in the absence of a professional – do not understand the standards or fails to submit images properly.

A common occurrence is when a person does not obtain a proper Vision ID and therefore cannot create a proper VIRIN. DIMOC often sees fraudulent Vision IDs such as XX111, AB123 or some other combination that clearly is not actually assigned to a person. This can, and does, create confusion in terms of properly tracking archived imagery.

Less common, especially with system-required fields at the DVIDS ingest point, are images with little or no supporting data; it is hard to capture context for these images and even harder to trace the source.

Since the initial recognition of the need for a unified written style for captions and metadata nearly two decades ago, DIMOC has continued to improve the value of its holdings. As asset management systems continue to improve, DIMOC has embarked on upon a new method of categorizing a very large collection of complex data.

Focus on Keywording

DIMOC's long-established and thorough captioning process has resulted in an abundance of descriptive metadata. With a new asset management system, DIMOC could finally fully consider using this descriptive information to increase the accessibility of its collection. DIMOC assessed all its metadata within the new asset management system, and began to consider implementing a controlled vocabulary to assist with workflow, accessibility, user-experience, and search and retrieval functionality.

Controlled Vocabulary Problem

Most of the metadata fields within DIMOC's schema have a controlled and finite list of terms for their entries [5]. While these fields, such as Combatant Command, have easy-to-build lists, the problem was not related to spelling consistencies. Though such errors existed, the problem was a lack of standards and consistency of the written word(s).

As *Introduction to Controlled Vocabularies* by The J. Paul Getty Research Institute states, "controlled fields contain data values drawn from controlled terms and are formatted to allow for successful retrieval. Free-text fields communicate nuance, uncertainty, and ambiguity to end users," yet DIMOC had long used a keywords field with no rigor or standard [6]. This open text field held misspellings or terms with and without punctuation, nicknames, abbreviations and terms that were bulk-edited/added and not necessarily applying to an individual image (within a series submitted from a larger event). The new methodology would have to correct these past inconsistencies while offering an arrangement that followed U.S. Department of Defense (DoD) structure, and incorporates the nuances and high-technical level of such a diverse language.

DIMOC's archive of visual information includes a collection of associated metadata for each DoD and Military Service still, motion, graphic and audio asset. Because each of the services has separate network capabilities and independent approaches to documenting and creating content, it would be thought that their independence also stretches to their imagery metadata. Yet within DIMOC's centralized system, a quick study of the metadata reveals redundancies in terminology across the services; however, it is not controlled. Without establishing a standard or controlled vocabulary to unite this content, inaccessibility or failed search results become the norm.

With a new asset management system and a mass digitization contract for legacy and born-digital imagery, DIMOC for the first time examined its accessibility issues with an industry-level system focused on the empowerment of metadata. The necessity to control the metadata and therefore unite like content became the critical focus to optimize search.

The definition of a controlled vocabulary, according to The Getty, is "an information tool that contains standardized words and phrases used to refer to ideas, physical characteristics, people, places, events, subject matter, and many other concepts. Controlled vocabularies allow for the categorization, indexing, and retrieval of information" [7].

Use of The Getty's *Introduction to Controlled Vocabularies* assisted considerations for implementing a controlled vocabulary and also for the application and software requirements, and for the improved search and retrieval functionality desired for a beneficial outcome. Although the book was written for the cultural heritage community, it was surprisingly easy to translate and apply to a federal government and military collection with technical nuances.

DIMOC's Requirements

The original purpose of DIMOC's controlled vocabulary was simply to be used within the asset management system and to automate some metadata editing. However, as knowledge was gained about the power of controlled vocabularies throughout DIMOC's community, multiple uses became visible and desired. While the data inconsistencies were problematic, even before the new asset management system, DIMOC long understood its strength lied within its metadata. Due to documentation requirements for U.S. Federal Government Records Management [8], DIMOC's data collection for each image goes beyond the basic metadata schema standards, such as the ubiquitous International Press Telecommunications Council (IPTC) metadata schema [9]. DIMOC's new asset management system offered the potential for greater use of its metadata, and encouraged standardization in a structured and weighted approach to its search and discovery abilities.

Selection of the Software Application

The vocabulary creators were discouraged to not find one dominating source for military terminology – even from a foreign country's cultural heritage institutions. DIMOC personnel creating the controlled vocabulary were doing so without precedent, leaving them wary. However, the secondary reception by the non-governmental archives and other institutions is just as concerning. These organizations that exist in a non-governmental environment are likely to see potential gaps, and the natural, albeit unknown, constraints created to form the structure and use of the vocabulary easier. These reservations established a requirement for a simplistic software selection – one that offered an easy export of both the terms and structure in cautious preparation for a pivot to another software should the situation present itself.

While the structure of the authority was considered, the selected software largely led DIMOC to a rigid approach, especially after generic hierarchical terms were considered the top-level (e.g. people, places, things, activities and events). However, The Getty states that compliance with the International Organization for Standardization (ISO) or National Information Standards Organization (NISO) standards should be followed [10]. Most controlled vocabulary software available for U.S. Government procurement inherently offers compliance with ISO and NISO standards, but the flexibility needed for practical organization drove DIMOC's product selection. Unfortunately, DIMOC found little establishment of a controlled vocabulary for military terminology. While consideration was paid to a number of templates, and even vocabularies that contained military terms, the organization and design was based upon the software's capabilities and on the visual clarity for the end user during the search.

Ultimately, the selection of an application focused on the display and user interface for general public consumers as well as DoD users. In other words, the visualization of DIMOC's vocabulary was central to its proper use and more importantly critical to its employment in search, discovery and accessibility. DIMOC personnel studied industry vocabularies and taxonomies of all types and across various subjects. The ones most tangible to the user had strengths in their visualization tools, leading to an easier translation to search, and therefore, greater discoverability, regardless of the level of technicality of the terms or language as whole.

While The Getty's *Introduction to Controlled Vocabularies* was used as a reference, it was coincidental that DIMOC's primary requirement emerged as the visual depiction, as it is stated, "Designing a good display is critical. Catalogers' or other users' willingness and ability to use the vocabulary are dependent upon how well they can understand and find terms. There are several types of possible displays, ranging from simple alphabetical listings to complex graphical displays. It is often desirable to provide multiple views of the vocabulary, including hierarchical displays, full record displays, and search results displays. Various methods of display, typography, capitalization, sorting, and arrangement of the data on the page or screen can be used to make terms easy to find and understand [11]."

Collection of Terminology

The collection of the terminology came directly from the U.S. military services and a list of unique values from DIMOC's asset management system's keyword metadata field. The services all maintain a fact file sheet on their equipment, vehicles, career fields, offices and overall organization. Some of these fact file sheets are more robust than others. In addition, DIMOC requested vocabulary lists from the service flagship webmasters (e.g. www.navy.mil or www.army.mil, etc.) [12].

The U.S. Coast Guard (USCG), while considered a military service, is part of the Department of Homeland Security (DHS), not the DoD. However, USCG imagery is collected by DIMOC via a formal agreement with the USCG's records management and with the U.S. National Archives and Records Administration. Because this content flows into the asset management system, the decision committed DIMOC to include USCG terminology. The combining of similar terms from the USCG's fact file sheet caused some initial authoritative confusion for the creator, specifically with the U.S. Navy's terminology. In the end, the Coast Guard vocabulary was added in similar arrangement to the other services.

Collection of terms such as foreign cities and countries, and bodies of water was completed from other government sources such

as the U.S. Department of State and the National Geospatial Agency respectively. The remaining boundary for how much DoD terminology to include was left to what was within the various service fact file sheets; additional terms will be added as the collection grows.

Saturation Sensitivity

Introduction to Controlled Vocabularies' states, "No vocabulary can contain all terminology. Boundaries for the vocabulary should be set, and the realm of knowledge encompassed should be precisely defined." [10] This axiom became a constant reminder of the potential oversaturation point. Maintaining this awareness stretched not only to the scope of the vocabulary – what would it contain – but also what relationships would be utilized. Overdoing either the volume of terms and/or an abundance of classes and/or the relationships constructed would effectively cancel out the implementation of a controlled vocabulary. In short, too much or oversaturation of either the terms or consistency organization is a detriment and counteracts the purpose of a controlled vocabulary.

The scope or reach of the vocabulary is made by the imagery itself – when an image contains a term not present in the vocabulary, its addition is made by manual validation (and via a report generated weekly by the asset management system for all unique terms). This is not to say that the vocabulary was created through an index and report of all the unique terms. Instead, a decision was made from that point forward to collect the new unique terms, not to go back through the entire collection of DIMOC's imagery and document the unique terms. That said, upon migration to a new asset management system, the keyword fields for all the assets were erased. The decision to move forward from this point with the standard keywords would be consistent – and nothing was critically missing given the descriptive quality and characteristics of the caption/description as a substitute for those assets now missing keywords.

Structure of the Vocabulary

The Getty has good suggestions, but for DIMOC's reactive state considering the "construction methods, plans for maintenance, desired structure, types of relationships, display formats, and policies regarding compound terms, true synonymy, and types of acceptable warrant" became out of reach given time constraints. After procurement, DIMOC did take the time to follow one of the more critical steps according to The Getty: "Constructing a rich and complex controlled vocabulary or authority is a time-consuming and labor-intensive process. However, the benefits are worth the cost, because the resulting vocabulary helps to ensure consistency in indexing and facilitates successful retrieval. It also saves labor, because catalogers do not have to repeatedly record the same information [13]."

The creation of the vocabulary's structure was not an easy endeavor. The DoD is a very complex organization, and therefore the first reaction was to create the structure to mimic the administrative and organizational hierarchy of the various DoD agencies. Specifically this organizational structure would place the imagery collection from the services at the top of the hierarchy. This approach initially created a difficult visual effect that was flawed with duplicative terms. While the terms themselves were only within the vocabulary system once, their repetition across the services' structure created a visual depiction full of crossed lines within the radial map view (figure 2), and within a hierarchical table (figure 3) the NT or narrower terms (children) were in too many object classes – U.S. military services – for there to be a clean top-down visual. For

illustrative purposes an overly simplistic example can be used: the Army and Marine Corps both use the same type of tank, or the Navy and the Air Force have the same career field. These lateral interactions become too chaotic and upon a visual appearance too difficult to follow. This structural design was flawed even for internal users with an understanding of the terminology.

The semantic mapping of terms underneath the services at the top of the hierarchy became too complicated for the visual depiction desired for users both internal and external. However, this flawed design of the services at the top of the vocabulary structure actually proved to be the gateway to the answer. In Figure 3, the complexity can be seen within the lateral duplication across the services' columns. If we draw attention to the rows, the same way a pivot table would showcase an alternate perspective, the row headings: People, Places, Things, Activities and Events, hold greater importance, and remove the duplication within the services by pushing the services to a categorized or thematic purpose. The use of these generic and even basic top-level terms turned a complicated organization into a clearly understood organizational structure, for even those who do not know military terminology can accept terms found in these headings. The Getty states, the respective "genus/species, or generic relationship, is the most common relationship in...taxonomies because it is applicable to a wide range of topics." [14]

U.S. military services' names (Army, Navy, etc.) within the controlled vocabulary will still be used, but as tags or per the software's designation as a category to individual terms. This functionality permitted collections to be built around the service names, so one can find all the Navy aircraft while still having a search filter for the generic "aircraft" lead to other services' aircraft (or to other degrees within the use of aircraft). The power of both systems is implied within this example: the asset management system's capabilities in a four-form (optional) search (e.g. open-text, keywords, filters, and refine search) is empowered with the integration of the controlled vocabulary's relationships to include the categorical tagging tool. Aligning what had seemed to be a difficult terminology structure with the search and retrieval function had been effectively achieved with the integration of these two systems.

Relationships

Without relationships between terms, no value to a controlled vocabulary exists, leaving the vocabulary without context or connection. DIMOC's vocabulary centers on the BT, or broader term, and NT, or narrow term. The hierarchical organization and structure of people, places and things, configured in the semantic mapping made the creation of these parent-child relationships obvious. The use of associative or "related to" relationships became the form for connecting terms across the classes.

All terms have a singular location within the vocabulary overall. The decision not to use a Multi-Narrow or Broad term relationship(s) was made based on the software's performance and visual depiction of these multiple relationship values. When implemented, the multi-relationships would be held at the top of the hierarchy (expansion) view, and therefore implied an inaccurate placement/location of the term within the generic classification of people, places, things, activities and events. One of the first internal processes for DIMOC became the consideration of a term as it is in a standalone assessment for addition to the vocabulary. This caused the associative or "related to" relationship utilization to fill in for the lack of multi-relationships. Understanding that "associative relationships are always reciprocal" was crucial and fortunately required the constant consideration of the search results. A career field that maintains a specific type of equipment is related to the equipment itself, therefore searching the

equipment (classified within "Things") will offer the related term(s) within the "like results," such as those within career fields.

The saturation awareness was critical at this point in the development. Selection of where the "related to" relationship occurred with the various hierarchical levels was mapped out as well and ultimately determined to be logical at the highest level. Therefore, individual ships are not "related to" their personnel or equipment, instead the relationship occurs at the class or type of ship. For instance, aircraft carrier is a type of ship, wherein the USS Carl Vinson (CVN 70) is an individually named aircraft carrier, and a part of the Nimitz-class; the "related to" mapping occurs at the type level.

Visually, the related term relationships offered accurate arrangement of the terms, unlike the multi-relationships display. DIMOC does not use multiple vocabularies directly, though it is a matter of interpretation. The Navy has their own terminology and preferences, and the other services have their own preferences too. It was determined that when a service-specific term, was entered within the vocabulary it would be maintained as service-specific but with a related term relationship to the other services' terms of similar meaning, use or purpose. The use of categories for the service names is the driving influence, but consistency of using a category versus maintaining a term organized as service-specific is yet to be seen, and will likely result in internal process policy.

Visual Display

One of DIMOC's biggest complaints from users and customers has been its search functionality, and implementation of a controlled vocabulary was designed specifically to respond and correct accessibility failure. The visual depiction is a priority to DIMOC when communicating to our users and customers. The complexity of military terminology and the nuances therein can be a difficult communication task among members of the different services, and it is even more ridiculous to expect general public users to easily grasp. DIMOC believes the use of a visual illustration of the vocabulary at large or in smaller organizational groupings, would offer an educational component for our customers and users at large. Figure 4 provides examples of the four visual tools offered to both the customers and internal users by the software DIMOC selected.

Giving users the ability to understand the hierarchical organization of our vocabulary upon their searching, the "like results" or "refine search" functionality becomes critical, as more online websites direct users to these added tools. An organizational structure of the vocabulary with the services at the top of the hierarchy caused too many lateral interactions to promote an understanding of the terminology to then expect the user to extrapolate how to achieve an accurate and relevant search.

Workflow

The maintenance of the controlled vocabulary will be conducted by the DIMOC Lead Writer-Editor, Archivists and Media Managers, all of whom closely work with the editing staff. However, the Archivists maintain the lead for the finding aids or collections management of the content, and therefore, are the enablers of the vocabulary in the asset management system.

The internal portal of the controlled vocabulary software offers an approval process for new terms, but the use of the unique terms weekly report from the asset management system is a single-person function with collaboration where discrepancies may exist. Regulating the organization of new terms is determined by the Archivists in consultation with the Lead Writer-Editor when an issue may arise, though this has yet to happen given the rigidity of the hierarchical structure of the fact file sheets; the decision is largely

made inherently with the addition of a term by the services. DIMOC requires a legitimate source to be cited upon approval of a new term, and within each term's definition exists a "resource" field to record the proper citation. This mandatory requirement largely drives the acceptance of a new term with little subjectivity involved. A new term may be approved if 1) used in an image, therefore the source can be ascertained by the editor during the validation process, and 2) if an addition is made to the services' fact file sheets. The Archivists and the writer-editor regularly check the latter. The flexibility within the system(s) and the workflow is key to gaining traction, comprehension and overall establishment of the controlled vocabulary as an integral part of DIMOC's business processes.

Automation

Automation becomes essential with the use of DIMOC's controlled vocabulary. Upon ingestion, the image's embedded metadata is written to the system's corresponding fields. The description as written by the photographer/videographer is indexed and terms within are recognized as matches to the controlled vocabulary. The term list for the high-level "things" is weighted higher than actions, events and people, respectively. Terms matching within the "things" list will be added as the first keyword, and if necessary a second keyword for things can be derived as well. The first keyword is always the primary subject of the image – and within the DIMOC collection this usually is an item from things. A majority of our users search for an object or thing. The second keyword, if not from the things list, will be the activity or event within the image, and the third keyword will be from the people list. During the lifecycle of an image, an editor will validate this automated process in addition to adding a fourth and fifth keyword – one of which is designed to be a concept.

Editors are becoming aware of the structure of the terms. More importantly, they are becoming aware of the impact of these essence keywords within the asset management system. These first three keywords are weighted more than the other terminology in the same metadata field. The first keyword is the primary feature of the image and holds more weight within the search results, and so do the second and third keywords.

These simplistic uses of three primary keywords, along with the "Description" or caption metadata field are the priority fields for an open text search. The keywords beyond the top three are used to complete the refine search results and the filters of like results. Experienced users who understand the identification numbering system or DIMOC-specific metadata have the liberty to search in these forms as well, but the focus is on the descriptive data available.

Search and Retrieval

Another important assessment from The Getty: "In retrieval, users do not always know what a person, place, or thing is called. Non-expert [*sic*] users often do not know the term used by a specialist to index a work... Expert users will know the specialized terms for a work, but different specialists may use different terms to refer to the same person, place, or thing" [15].

All users have different knowledge bases when it comes to military terminology. And, of course, the reasons for searching all vary greatly. Attempting to be all-inclusive would lead to failure in regard to the spectrum of users, just as there was saturation sensitivity for the volume of terms and organizational structure of the vocabulary. DIMOC decided to stick with what it knew and the data to support it. The terms collected became the authoritative list to include synonyms, variants and related terms, which were individually added, and/or relationships built based upon the hard

information from the data at hand. Certainly, there are gaps in common language, acronyms and additional synonyms among the unforeseen as well, and DIMOC's asset management system will assist in bridging these gaps. A report can be generated for all system searches and this data becomes extremely helpful when a search fails.

Users will be guided to the open-text box search with primary search run through two descriptive metadata fields: keywords and description. The users may then use the filters, refine search open-text box or like-results keywords, which is driven by the fourth and subsequent keywords. Metadata that holds more value for internal users' searches has the same controlled vocabulary applied for the descriptive metadata fields, though internal users are more likely to use the definitive list metadata fields versus the large volume of descriptive terms metadata fields. These metadata fields with definitive lists include the Combatant Commands field, for which only nine authorized terms exists (abbreviations included), or three primary identification number metadata fields.

The advanced search functionality of the asset management system will be expanded with the capability to search all 120 metadata fields DIMOC maintains using Boolean operators, and other qualifiers such as "contains" or "does not contain" for added flexibility. DIMOC anticipates the advanced search to be used primarily by its own personnel, though it will be available to all user types. All the descriptive metadata fields are indexed and used in the filters of the asset management system. These descriptive fields go beyond the caption and keywords field searchable in the open-text box. These fields include: Service Shown, Command Shown, Service Documenting, Year, Era, (Military) Operation and Exercise, and the three geographic fields (Base, State, Country). The geographic metadata fields are also automatically populated based upon the description. This descriptive metadata is indexed and used within the filters of the asset management system.

DIMOC's primary requirement for the controlled vocabulary is to educate or at least offer guidance to the user via visual depiction of the terminology or via search and retrieval functionality. The Getty states that how users are guided or led to use the vocabulary matters greatly in that "end users may be guided in their searches by showing them the expert terminology from which to choose, in a process known as *user intervention* or *mediation*" [16]. Displaying vocabulary in all four search tools offered by DIMOC's asset management system meets this passive communication technique to the end user, as well as the available visualization tools.

Potential Partnering Opportunities

Given the vocabulary's organization within the generic, an unintended byproduct of the DIMOC's DoD Controlled Vocabulary is that it can be easily be cross-referenced to the Library of Congress Subject Headings (LCSH), however, the variety of nuances within the technical names of the terms, as well as their use, does not guarantee a direct relationship to the LCSH. As the first known controlled vocabulary of the Department's visual imagery, DIMOC is hopeful to establish its vocabulary as an appending standard to LCSH, among others in the next year.

DIMOC's DoD Controlled Vocabulary has been requested by a number of services' webmasters for both internal and external websites. With DIMOC's collection of the services' fact file sheets, the already promulgated terms have been centrally collected in an accessible (e.g. API request) organization and structure, and with the added benefit to the requestor that it will be maintained and kept current.

DIMOC's previous lack of participation in the archival community is peculiar given its long history as an archive. Notably,

in the last three years, DIMOC has taken a number of steps to increase functioning as a more public archive (than a mere internal-only governmental archive): mass digitization contract, new asset management system, refreshed workflow, personnel training to industry and archive standards, and efforts at transparency, an example being the numerous papers written for IS&T Archiving 2016 Conference. Higher management, and a relevant larger agency strategic plan support DIMOC's recognition as an archive belonging to a larger community outside the federal government. DIMOC can only assume that potential partnerships, whether formal or informal, will occur as an outcome of its efforts in metadata management, accessibility, search and retrieval, media digitization, or any other number of functions DIMOC performs.

Conclusion

Maintaining standards for both the overall captioning process from the photographers – submitters – and within the DIMOC editing processes will only continue to benefit the imagery collection as its size and access grows. With the digitization of DIMOC's archive, the metadata standards and controlled vocabulary become even more critical for accessibility. The use of metrics to monitor search functionality success will provide feedback, revealing the performance of these standards. DIMOC's metadata management will continue to evolve, but the ultimate goal of increasing the discoverability of the content will remain the same.

References

- [1] U.S. Department of Defense Captioning Style Guide, 17 October 2014, <http://www.dimoc.mil/documents/styleGuide/StyleGuide-141017.pdf>
- [2] DF-SC-16-12846, DIMOC holdings
- [3] U.S. Department of Defense Instruction 5040.02, *Visual Information*, 27 October 2011, <http://www.dtic.mil/whs/directives/corres/pdf/504002p.pdf>
- [4] The Associated Press Stylebook, Current Edition
- [5] DIMOC's DoD VI Controlled Vocabulary: <http://dimoc.synaptica.net/pvs/browse.asp>
- [6] Harping & Baca, Introduction to Controlled Vocabularies, Controlled Fields vs. Free-Text Fields http://www.getty.edu/research/publications/electronic_publications/intro_controlled_vocab/retrieval.html
- [7] Harping, Patricia & Baca, Murtha, Introduction to Controlled Vocabularies: Terminology for Art, Architecture, and Other Cultural Works, *2010 J. Paul Getty Trust, Controlled Vocabularies in Context*: http://www.getty.edu/research/publications/electronic_publications/intro_controlled_vocab/context.html
- [8] Electronic Code of Federal Regulations, US Government Printing Office: <http://www.ecfr.gov/cgi-bin/text-idx?node=36:3.0.10.2.10&>
- [9] International Press Telecommunications Council (IPTC) metadata schema: <https://iptc.org/standards/photo-metadata/>
- [10] Local Authorities http://www.getty.edu/research/publications/electronic_publications/intro_controlled_vocab/local.html
- 10 - Harping & Baca, Introduction to Controlled Vocabularies, Characteristics of Displays http://www.getty.edu/research/publications/electronic_publications/intro_controlled_vocab/constructing.html
- [11] Harping & Baca, Introduction to Controlled Vocabularies, Characteristics of Displays http://www.getty.edu/research/publications/electronic_publications/intro_controlled_vocab/constructing.html
- [12] Fact File Sheets: US Army: <http://www.peosoldier.army.mil/portfolio/#1>; US Navy: <http://www.navy.mil/navydata/fact.asp>; US Air Force: <http://www.af.mil/AboutUs/FactSheets.aspx>; US Marine Corps Equipment: <http://www.marines.com/operating-forces/equipment>
- [13] Harping & Baca, Introduction to Controlled Vocabularies, General Criteria for the Vocabulary http://www.getty.edu/research/publications/electronic_publications/intro_controlled_vocab/constructing.html
- [13] Harping & Baca, Introduction to Controlled Vocabularies, Semantic Mapping http://www.getty.edu/research/publications/electronic_publications/intro_controlled_vocab/multiple.html
- [14] Harping & Baca, Introduction to Controlled Vocabularies, Relationships in Controlled Vocabularies http://www.getty.edu/research/publications/electronic_publications/intro_controlled_vocab/relationships.html
- [15] Harping & Baca, Introduction to Controlled Vocabularies, Retrieval Using Controlled Vocabularies http://www.getty.edu/research/publications/electronic_publications/intro_controlled_vocab/retrieval.html
- [16] Harping & Baca, Introduction to Controlled Vocabularies, User Intervention or Behind the Scenes http://www.getty.edu/research/publications/electronic_publications/intro_controlled_vocab/retrieval.html

Author Biography

Thomas M. Ruyle joined DIMOC as the Lead Writer-Editor in March 2012. In this role, he works with Department of Defense personnel to ensure accuracy and completeness of captions and metadata associated with DIMOC holdings. In between two stints in the U.S. Army, Ruyle earned a B.A. in History from the University of North Carolina at Wilmington in 2001. He was also a copy editor at several newspapers, including Stars and Stripes.

Julia Hickey received her master's in Public History from the University of Wisconsin –Milwaukee. She is currently an archivist for visual information for the Department of Defense within the Defense Imagery Management Operations Center, Defense Media Activity, Fort Meade, Maryland.

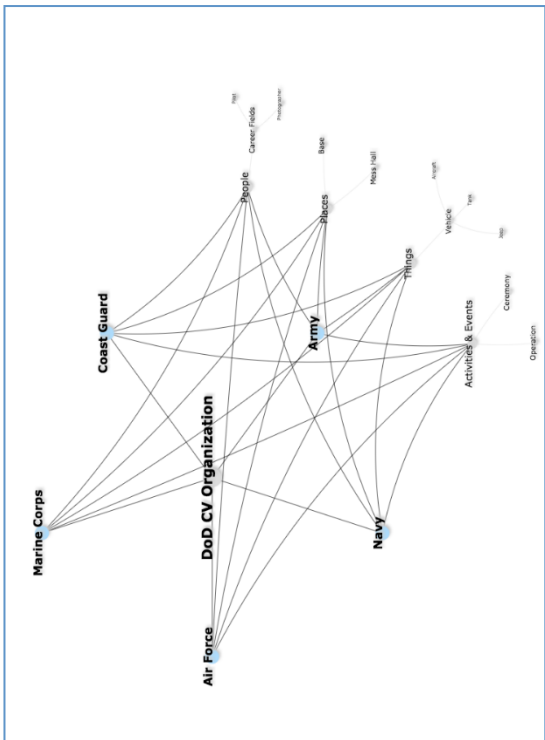


Figure 2. Radial Map, DoD VI Controlled Vocabulary

	Army	Navy	Air Force	Marine Corps	Coast Guard
People	Career Fields	Career Fields	Career Fields	Career Fields	Career Fields
	Photographer Pilot	Photographer Pilot	Photographer Pilot	Photographer Pilot	Photographer Pilot
Places	Base	Base	Base	Base	Base
	Mess Hall	Mess Hall	Mess Hall	Mess Hall	Mess Hall
Things	Vehicles	Vehicles	Vehicles	Vehicles	Vehicles
	Aircraft	Aircraft	Aircraft	Aircraft	Aircraft
	Jeep	Jeep	Jeep	Jeep	Jeep
	Tank	Tank	Tank	Tank	Tank
Activities & Events	Equipment	Equipment	Equipment	Equipment	Equipment
	Helmet	Helmet	Helmet	Helmet	Helmet
	Weapon	Weapon	Weapon	Weapon	Weapon
	Missile	Missile	Missile	Missile	Missile
	Ceremony	Ceremony	Ceremony	Ceremony	Ceremony
	Operation	Operation	Operation	Operation	Operation

Figure 3. Hiearchical table, semantic mapping example.

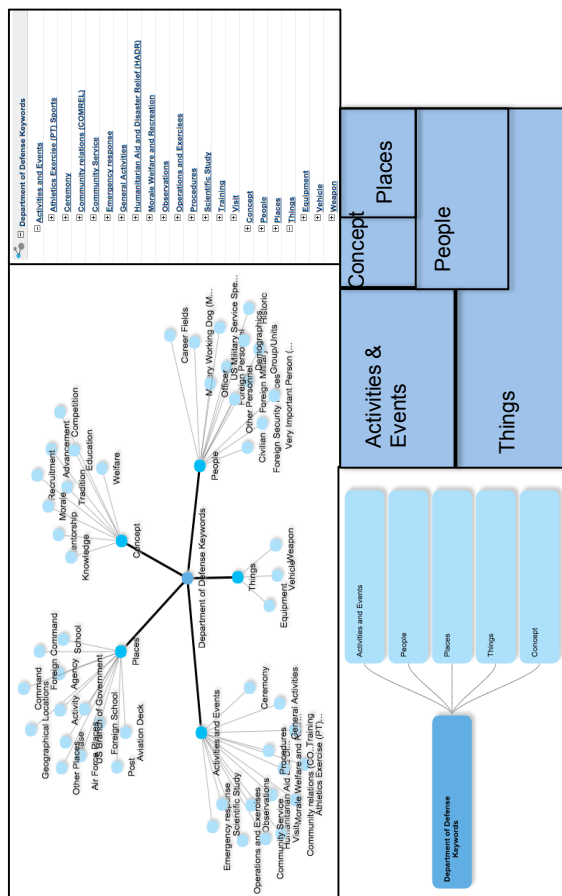


Figure 4. Four different visual views for the DoD VI Controlled Vocabulary. From left to right: Radial Map, Hierarchical Tree Browse, Tree Map, and Word Tree (illustration).