Data Analytic Considerations for Audio, Video, and Simulation Trace Data: Enabling Decisional Advantage

Ms. Jennifer Winner
711th Human Performance Wing
WPAFB, Ohio
Jennifer.winner.1@us.af.mil

Mr. Cameron Roudebush
Aptima
Dayton, Ohio
Cameron.roudebush.ctr@us.af.mil

Dr. Randall Spain
US Army DEVCOM Soldier Center
Orlando, Florida
Randall.d.spain.civ@army.mil

Mr. Ian Davis
BAE Systems, Inc.
Dayton, Ohio
Ian.davis.ctr.18@us.af.mil

Ms. Katelyn Kay, Dr. Kent Etherton
711th Human Performance Wing
WPAFB, Ohio
Katelyn.kay.1@us.af.mil, kent.etherton.1@us.af.mil

Ms. Erin Baker
Naval Air Warfare Center, Training Systems Division
Orlando, FL
Erin.a.baker2.civ@us.navy.mil

Dr. F. Eric Robinson
Naval Medical Research Unit – Dayton
WPAFB, Ohio
Frank.robinson.5@us.af.mil

Dr. Benjamin Goldberg
U.S. Army DEVCOM Soldier Center - STTC
Orlando, FL
Benjamin.s.goldberg.civ@army.mil

ABSTRACT

Data standards govern how digital data are formatted, organized, and stored to facilitate later use. The training value gained through the application of data standards has long-been acknowledged Standards have enabled real-time performance feedback and rigorous training research, and the application of standards has poised the operational training community to benefit greatly during the current explosion of artificial intelligence (AI) capabilities. Doing so, however, assumes that training systems are built with the capability to support seamless data manipulation and export, which analyses have shown is not the case (NAWCTSD & Katmai, 2023). Data trapped within a training management or after-action review system negates the potential of current computational advances. Complexities in the requirements process and coordination of training system requirements for acquisition, including cross-service collaboration, are acknowledged across the services (Marler et al., 2021; NAWCTSD & Katmai, 2023). Organizations must proactively identify data requirements during the design and development of simulation-based training and after-action review systems to realize the full value of their digital training data. Program managers considering open systems architecture and various data strategies have resources to guide them (e.g., Defense Acquisition University, 2013; Guertin & Hurt, 2013), however, these resources fall short of articulating needs for specific data types and analyses goals to support data-driven learning analytics. This paper addresses this gap. Specifically, we discuss how audio, video, simulation trace, and other multimodal data should be collected and formatted to support training analytics with emerging AI tools and techniques. Illustrated in the context of military medical training, these considerations are applicable in other domains. These recommendations can be leveraged by program managers looking to avoid roadblocks preventing efficient and effective use of their individual and team-level human performance data to inform training and operational decisions.

ABOUT THE AUTHORS

Ms. Jennifer Winner is a Senior Research Psychologist at the Air Force Research Laboratory where she conducts human effectiveness research for the 711th Human Performance Wing, Human Effectiveness Directorate, Warfighter Interactions and Readiness Division's Operational Learning Sciences Branch (711HPW/RHWO). She received her M.S. in Applied Psychology from Arizona State University in 2007, her M.B.A. from Wright State University in 2017, and is currently completing a Ph.D. in Integrated Systems Engineering from The Ohio State University (anticipated graduation in 2026).

Mr. Cameron Roudebush is a Senior Software Engineer at Aptima, Inc. He received his Bachelor of Science in Computer Science from Wright State University in 2020.

Dr. Randall Spain is a Research Scientist at the U.S. Army Combat Capability Development Command – Soldier Center. He received his Ph.D. in Human Factors Psychology and M.S. in Experimental Psychology from Old Dominion University. His research focuses on the design and evaluation of adaptive and intelligent training systems.

Mr. Ian Davis is a Software Engineer and Technical Lead at BAE Systems, Inc. He received his Bachelor of Science in Computer Science from Wright State University in 2019.

Ms. Katelyn Kay is a Research Psychologist with the Air Force Research Laboratory, 711th Human Performance Wing. She received her M.S. in Human Factors from Embry Riddle Aeronautical University in 2022.

Dr. Kent Etherton is a Research Psychologist with the Air Force Research Laboratory, 711th Human Performance Wing. He received his Ph.D. in Human Factors and Industrial/Organizational Psychology from Wright State University in 2021. His research focuses on building training tools to enable rapid skill acquisition and identifying data needs to evaluate simulator efficacy.

Ms. Erin Baker is a Senior Research Psychologist for the Naval Air Warfare Center Training Systems Division with over a decade of training and education experience in the medical space. Mrs. Baker received her M.S. in Industrial/Organizational Psychology in 2011 and her B.S. in Psychology from the University of Central Florida in 2007.

Dr. F. Eric Robinson is a research psychologist in the Naval Aeromedical Research Laboratory at the Naval Medical Research Unit Dayton. He earned his PhD in human factors psychology from Wright State University in 2017 with an emphasis on expert reasoning and environmental influences on decision making.

Dr. Benjamin Goldberg is a Senior Research Scientist at the U.S. Army Combat Capability Development Command – Soldier Center. His research focuses on adaptive experiential learning with an emphasis on simulation-based environments and leveraging Artificial Intelligence to create personalized experiences. Dr. Goldberg holds a Ph.D. in Modeling & Simulation from the University of Central Florida and is well published across several high-impact journals and proceedings, including IEEE Transactions of Learning Technologies, the Journal of Artificial Intelligence in Education, and Computers in Human Behavior.

Data Analytic Considerations for Audio, Video, and Simulation Trace Data: Enabling Decisional Advantage

Ms. Jennifer Winner
711th Human Performance Wing
WPAFB, Ohio
Jennifer.winner.1@us.af.mil

Mr. Cameron Roudebush
Aptima
Dayton, Ohio
Cameron.roudebush.ctr@us.af.mil

Dr. Randall Spain
US Army DEVCOM Soldier Center
Orlando, Florida
Randall.d.spain.civ@army.mil

Mr. Ian Davis
BAE Systems, Inc.
Dayton, Ohio
Ian.davis.ctr.18@us.af.mil

Ms. Katelyn Kay, Dr. Kent Etherton
711th Human Performance Wing
WPAFB, Ohio
Katelyn.kay.1@us.af.mil, kent.etherton.1@us.af.mil

Ms. Erin Baker
Naval Air Warfare Center, Training Systems Division
Orlando, FL
Erin.a.baker2.civ@us.navy.mil

Dr. F. Eric Robinson
Naval Medical Research Unit – Dayton
WPAFB, Ohio
Frank.robinson.5@us.af.mil

Dr. Benjamin Goldberg
U.S. Army DEVCOM Soldier Center - STTC
Orlando, FL
Benjamin.s.goldberg.civ@army.mil

INTRODUCTION

Data are of great strategic importance for the U.S. Department of Defense (DoD). The DoD strategy entitled "Data, Analytics, and Artificial Intelligence Adoption" opens with an emphasis on the criticality of data leverage. It states, "The latest advancements in data, analytics, and artificial intelligence (AI) technologies enable leaders to make better decisions faster, from the boardroom to the battlefield. Therefore, accelerating the adoption of these technologies presents an unprecedented opportunity to equip leaders at all levels of the Department with the data they need, and harness the full potential of the decision-making power of our people" (Department of Defense, 2023, p. 3). Without usable data, the advantages afforded by these emerging technologies will equate to missed opportunities. In this paper, we offer recommendations aimed at maximizing data availability with specific focus on human performance data in the form of audio, video, and simulator trace data collected in the context of military training environments.

Decades of research on integrated (networked) training and distributed mission operations have highlighted the advantages afforded by data sharing standards, which include everything from increased fidelity, opportunities for integrated teamwork and deliberate skills practice, enhanced performance assessment, and feedback for after-action review (e.g., Bell, 1999; Schreiber, 2013). Data standards govern how digital information is formatted, organized, and stored to facilitate later use, and standards are critical enablers for realizing the promised training value (Schreiber, 2013). Although the benefits are well-known (e.g., Hernandez, et al., 2022; NAWCTSD & Cole Engineering, 2021), these best practices have failed to be fully leveraged in some contexts. Recent observations in the medical modeling, simulation, and training context have revealed opportunities to increase capabilities to support data manipulation and export (NAWCTSD & Katmai, 2023). This lack of interoperability has resulted in lost opportunities for data export and data sharing, which has reduced capabilities for readiness validation and analytics on emerging concepts of operation. Our direct observations of simulation-based data capture in such settings have revealed that the lack of interoperability reduces data access, including but not limited to the audio, video and simulation trace data. Making data available for analysis outside of the system in which data were produced is critical to future operations.

This fundamental requirement must be accounted for in future system design. Vendor solutions that fail to address data sharing standards, application programming interfaces (APIs), or data export capabilities essentially trap data within a training management or after-action review system, often void of context surrounding its capture. This negates the potential of current computational approaches and undermines the DoD strategy's focus on maximizing data leverage for decisional advantage. Specific to training and education, requirements for future systems must address data availability and export in support of learning analytics (e.g., Goodell & Kolodner, 2022; Krumm, Means, & Bienkowski, 2018). If training is to be predictive of future performance, researchers and trainers require sufficient data to defend claims of predictive validity.

Cross-service collaboration and complexities associated with the requirements process are extensive (e.g., Marler et al., 2021; NAWCTSD & Katmai, 2023). Organizations must proactively identify data requirements during the design and development of new systems to realize the full value of digital training data. Program managers considering open systems architecture and various data strategies have resources to guide them (e.g., Defense Acquisition University, 2013; Guertin & Hurt, 2013; Wydler, 2014), however, these resources fall short of articulating needs for specific data types and analyses goals, including support for data-driven learning analytics and human performance analytics. Our paper addresses this gap by leveraging experience and expertise in the areas of engineering, learning analytics, human factors, human performance, industrial organizational psychology, and intelligent tutoring. The recommendations offered in this paper were informed by our respective agencies' investments in human performance and training effectiveness research. Specifically, we discuss how audio, video, and simulation trace data must be captured and stored to support learning analytics using emerging AI tools and techniques.

In this paper we discuss data availability in our illustrative use case (military medicine) along with data repository considerations, and we organize our discussion around three main data types: audio, video, and simulation trace data. In Table 1, we introduce a set of questions and considerations distilled from the later discussion to give a glimpse at what is ahead. We encourage program managers to consider these items when developing requirements for modeling and simulation, training, and after-action-review tools. Specific recommendations relative to these topics can be found in the later sections.

Table 1. Questions and Considerations for Envisioned Systems.

Questions	Considerations
How will the system support data availability and export?	• Which data does the system produce? • Which data are critical for export (e.g., metadata for context)? • Define the real-time and post simulation data access needs. • Define the expertise requirements relevant for data export use cases. • Clarify the unique HIPAA, PII, or security requirements.
How will the system support synchronization?	• Identify meaningful timestamp or synchronization needs. • Define the role time plays in reconstruction of events and analyses.
How will the system make quality data available for analytic tools?	• Identify critical data formats that enable or hinder analytic approaches. • Evaluate open standards and application programming interface needs. • Identify the formats that result in a loss of data or specificity. • Are there relevant data post-processing hurdles? • Identify data provenance needs to enable apples-to-apples comparisons.
How will the system enable export to external stores?	• Clarify the unique IRB requirements associated with data repositories. • Identify metadata requirements necessary for consent tracking.

DATA AVAILABILITY IN MILITARY MEDICAL TRAINING

We have drawn on experience across military medical training and other training contexts to distill the recommendations made in this paper. Specifically, the considerations we discuss were observed across projects executed by the Air Force Research Laboratory, the Naval Air Warfare Center Training Systems Division, the Naval

Medical Research Unit Dayton, and the U.S. Army's DEVCOM Soldier Center. In this section, we summarize the current state of data availability and interoperability within a generalized use case, the military medical domain, as observed through these projects. Following this breakdown, we transition into recommendations for writing data-informed requirements to address the observed gaps in data availability relevant to human performance and learning analytics. Note that these recommendations are applicable across a variety of training domains (not just medical).

Joint Medical Training Readiness Tracking: A Survey

Within the military medical system, specifically within education and training, current systems that track force readiness were developed under a multitude of separate contracts in support of service, agency, or facility-specific requirements. Many systems require extensive manual input, resulting in limited mechanisms available to enable data integration. Given that existing data are currently stored and managed separately by the Department of Defense (DOD), Army, Navy, Air Force, and commercial corporations and due to lack of connectivity and interoperability, there are missed opportunities for analyses. The outcome is that multiple systems must be used to track training and document compliance, with no means of aggregating the data automatically (NAWCTSD & Cole Engineering, 2021).

Interoperability standards govern interoperability compliance, which enables compliance with information security, privacy, and cybersecurity requirements. In some use cases, these standards harmonize the sharing of data (e.g., courseware, competencies scenarios) among organizations and systems to meet the business needs of the larger healthcare enterprise. Although these organizations and systems are separated by geography and goals, this harmonization provides the same format and methodology across the industry to ensure the portability of data between organizations. In other use cases, these standards ensure data exchange is standardized between disparate Learning Management Systems (LMS) and Learning Record Stores (LRS) for effective information sharing (Walcutt & Schatz, 2019). Following a standardized format and methodology allows independent learning platforms to interoperate as though they were designed to do so. As a result, data would be able to flow freely, but securely, between these systems providing learners with the full breadth of available resources (NAWCTSD & Cole Engineering, 2021).

Interface standards implemented at the requirements stage aid in preventing delivery of "turnkey" systems where proprietary interfaces link components and result in high lifetime costs because the system is not optimized for the user's particular needs. Competitors will typically offer components that are superior to some of those in the turnkey system and price competition will not be a factor when system components need replacement. In such situations, system design can still be optimized. However, the cost of modifying physical and functional interfaces to allow components from different vendors to work together (i.e., to "interoperate") is usually prohibitive, and full functionality is often not obtained by reengineering proprietary interfaces (NAWCTSD & Cole Engineering, 2021).

In summary, the current state of our data infrastructure is ripe with opportunity for advancement. Along with the adoption of standards in the requirements generation process, there is a need to develop implementation best practices that align to associated data-analysis needs. In the remainder of this paper, we focus on common data types and discuss considerations that are critical in the context of human performance data and learning analytic needs.

RECOMMENDATIONS TO SUPPORT DATA ANALYTICS

In this section we outline several recommendations organized around data type and specific analytic goals. First, we address differences in data for tracking in the context of our illustrative use case, medical team training, which include distinctions between clinical encounters, training system utilization, and human performance outcomes. Second, we discuss the implications of data repositories and how those requirements extend well beyond hardware and software. Third, we outline recommendations to address data format needs for audio data captured in training environments. Within this area, we discuss data needs for content and flow-based measurement. Fourth, we outline recommendations to address video data formats for automated analysis of imagery collected within training environments. Fifth, we discuss simulation trace data that is needed to cross-correlate with many of the other data types in support of various analytic needs.

It is impossible to address all human performance data and metrics across the services given the limited space within this paper. Rather, we aim to address some common data types relevant for individual-level and team-level analyses. Although relevant training data are omitted from detailed discussion here (e.g., grade sheets, demographics, training completion), we believe that our sampled set achieve our goal of informing future system requirements in ways that account for data analytic considerations and increase the availability of usable, exportable data.

Encounter, Utilization, and Human Performance

Before diving into the topic, there are a couple of distinctions uniquely relevant in our illustrative context of the military medical training space. Those include the distinctions between clinical encounters, utilization, and human performance data. Clinical encounters are defined within the Defense Health Agency's procedural instructions, and they refer to instances of face-to-face interactions between providers and patients (Defense Health Agency, 2018). These encounters are distinct from measures related to simulation center and equipment utilization. Training system vendors may offer solutions providing equipment usage statistics. This may be helpful in evaluating equipment useful life remaining and estimating hardware needs across different sites. However, it is important to note that this data falls short of feeding analytics regarding human performance and the training efficacy of those sessions, trends to understand the rate of skill acquisition within learner audiences, trends and changes within a training pipeline or course over time and estimates for potential future concepts of operation based on past performance.

Ensuring the efficacy of training depends on the availability of appropriate data to assess claims of predictive validity. That is, it is impossible to empirically test the degree to which a warfighter's training performance is a valid predictor of how they will perform operationally if the appropriate data does not exist. The appropriateness of such data depends on the use case of interest, but across all situations, it is critical to operationalize warfighter performance along several human performance metrics. In-depth analysis is needed to identify the appropriate human performance measures to support readiness assessment and after-action review is a vital piece within simulation-based training and after-action review systems. Maximum data leverage, from the human performance perspective, includes not only simulation trace (event) data (i.e. "what happened in the scenario"), but also trainee or learner actions (i.e. "what did they do"), actions by the instructors and simulator operator, and the resulting outcomes (e.g., patient outcomes, changes in patient status). Lastly, it necessarily includes data provenance to know that one is comparing apples to apples down the road, as data is leveraged in the future for emerging needs. Possessing the necessary data to glean such insights allows researchers and trainers to assess the predictive validity of training simulations.

Data must be sampled, stored, and exported in workable formats to enable analyses. In this paper we present recommendations related to these aspects of human performance data as they are split across data types. For the data types we discuss, there are relevant research data protection considerations. Thus, we first discuss the role of data repositories toward this end.

Feeding Human Performance Data Repositories

In simulation-based training contexts, data repositories are more than the sum of the hardware and software required for storage. Repositories inherently require personnel and processes to manage the informed consent process, address the protection of data collected from human research subjects, and continuously direct data collection to ensure sufficient metadata are captured. Accordingly, a data repository serving learning analytics and research is something that needs to be continuously designed, managed, and maintained, as opposed to a solution purchased off-the-shelf.

Human performance data use in scientific research is regulated by Institutional Review Boards (IRBs) and Common Rule Code of Federal Regulations (Common Core, 2018), therefore a data repository intended to support future research must be supported by an overseeing IRB. Data repository approval requires documentation such as a protocol, a list of associated data managers, and informed consent documents. Data protections will differ by sensitivity of data. Protected Health Information (PHI) and Personally Identifiable Data (PII) require additional protections to ensure safety within the repository (Perazzo et al., 2019). For data that can be de-identified, gathering full consent (under IRB guidance) is an advantageous measure to allow for long-tail data reuse. Upfront efforts to obtain and manage consent, including specification for future data use, and can therefore benefit multiple future research goals. Adopting a multimodal data repository solution is strategic for data preservation, access, and reuse of big data. Exportable datasets are most effective when supported by context rich metadata (Greenberg et al., 2009; Martin et al., 2017; Trisovic et al., 2021).

Metadata should be used to connect various types of data pertaining to the same instance of collection. Contextual metadata includes participant identification, date of capture, and other variables that can be used to navigate mass data sets. Data provenance, or the origin of data, is a critical factor when designing and managing a data repository. For instance, recorded training scenarios must be tagged with dates of training and participant IDs to enable analytic connections to other types of relevant data such as subject demographics or student grade sheets. These contextual pieces of metadata must be included in mass data export to ensure accuracy in later analyses, and to support data management needs in instances where research data participants withdraw consent. Without the ability to connect metadata to relevant outputs, research data are less informative and require more intervention to untangle. In summary, it is critical to appropriately characterize the data management and protection responsibilities associated with data repositories. Having introduced some of the ways metadata plays a critical role in effective data use, we now turn to discussions and recommendations organized around data source.

Audio Data Recommendations

Audio data are a critical enabler for analysis of team performance in complex settings such as military medicine. The recommendations we present draw from existing industry practices and our experience developing advanced analytics, for which we leverage recordings of military medical teams capturing during simulation-based training. Our analytics involve calculation of scores depicting team behaviors such as adaptation, reorganization, and leadership. These calculations are based on derivations from the actual audio data, such as time-stamped transcriptions (what is said) and diarization (who said it) files segmented by individual speaker (Reynolds et al., 2005). The accuracy of the transcriptions and diarizations is critical for ensuring optimal audio capture. There are countless elements of audio recording that may impact data quality. Those critical to the transcription and diarization processes are explored here.

Our review of available IEEE standards revealed a stark lack of relevant guidance on audio capture and storage within the context of research analysis. IEEE 1857.8-2020 does articulate standards for audio, but strictly within the context of streaming it over a network connection (IEEE Standard for Second Generation Audio Coding, 2020). IEEE 3302-2022 describes systems to enhance audio for moving picture using artificial intelligence, such as emotion-enhanced speech, audio recording preservation, and speech restoration (IEEE Standard Adoption of Moving Picture, Audio and Data Coding by Artificial Intelligence (MPAI) Technical Specification Context-based Audio Enhanced (CAE) Version 1.4, 2023). IEEE 1857.2-2023 describes new audio coding algorithms for lossless audio compression, which may prove useful in the future, but provides little guidance on audio capture methodology (IEEE Standard for Advanced Audio Coding, 2023). None of these standards address the design choices we have seen threatening audio data access and quality in military medical training settings today.

Audio File Formats and Sizes

Common file formats for audio recordings include Waveform Audio File Format (WAV) and Moving Picture Experts Group (MPEG) Audio Layer-3 (MP3) (den Uijl et al., 2013; IASA Technical Committee, 2009). The WAV format is a lossless, uncompressed audio format that stores raw linear pulse code modulation (LPCM), which directly corresponds to the data output from most recording devices. Alternatively, an MP3 recording is a compressed audio format that aims to significantly reduce file size by carefully removing audio features from the original recording. The removal of audio features by compressed formats hurts the accuracy of the transcription and diarization process, and should be avoided (Ng et al., 2004). There are also lossless compressed audio formats, such as Free Lossless Audio Codec (FLAC) and MPEG4-Audio Lossless Coding Scheme (MPEG4-ALS). However, managing codecs and audio for these formats adds complexity to data processing as most transcription and diarization processes do not natively support these file formats, thereby requiring audio conversion. Thus, using the WAV audio file format would be the best practice, but FLAC or MPEG4-ALS with appropriate codecs and audio conversion tools could be used instead to maximize storage space while preventing any quality loss (Harada, Moriya, & Kamamoto, 2007).

The size of audio files will also be influenced by the audio sampling rate. The current standard sampling rate is 44.1kHz (Garcia et al., 2020), though most transcription and diarization processes sample audio at 16kHz, resampling input data as necessary. Thus, the minimum required sampling rate is 16kHz, with 44.1kHz as a best practice for maximum compatibility with modern audio software and tools.

The Distributed Interactive Simulation (DIS) standard is a major player in the training and simulation space. Reviewing the DIS Standards for radio and audio transmission, we see the supported audio formats: 8-bit μ-law, CVSD, ADPCM, 16-bit Linear PCM 2’s complement, 8-bit Linear PCM. Of these formats, only 16-bit Linear PCM 2’s complement should be used, as all others use compression or a limited bit depth that will hurt audio analysis. The standard also allows for a range of sample rates between 8kHz and 48kHz. Where possible, avoid a sampling rate below 16kHz to prevent transcription and diarization accuracy loss (IEEE Standard for Distributed Interactive Simulation–Application Protocols, 2012).

Audio Tracks

Audio recordings captured during training can either be captured using a single acoustic microphone positioned centrally in a room (i.e., single track) or by having a dedicated microphone worn by each participant (i.e., multi-track). When using only one microphone, the resultant audio file has one single track with speech from all participants together as one audio waveform. While this approach boasts easy setup and minimal user burden, it adds multiple pitfalls to the analysis process. Common transcription models do not perform well with multi-speaker, single waveform data, particularly in the presence of "step-ons," when the speech from two or more participants overlaps (Park et al., 2022). This can result in major transcription accuracy loss. Moreover, the need for diarization to effectively segment the audio file by speaker can further limit the validity of any subsequent analysis. Alternatively, the use of multi-track audio recordings completely removes the need for diarization, representing a substantial improvement. This reduces the negative impact that overlapping speech events on transcription accuracy, since only one speaker is represented in each recording track. Consequently, our recommendation is to utilize individual recording devices to capture separate audio tracks for each participant in an exercise, and store audio files with each track separated. In cases where a single-track audio file is required, audio manipulation software can achieve this by "mixing down" the multi-track source file into a mono audio file.

Audio Analysis

We have discussed the criticality of being able to successfully export data for analysis outside of the system in which they were captured. The audio analysis process is a generally domain-specific topic, meaning the analytic standards and comparisons may vary widely depending on research needs. The types of team-level communication measurement we are advancing analyzes audio data, which involves the preparation of transcripts and then implements algorithms to evaluate communication-based metrics (Harrison et al.,2023; Gorman et al., 2020; Gorman & Wiltshire, 2022). We use a secured data store dedicated to the audio recordings, transcriptions, and diarization data. Others conducting similar research may use data store systems in the form of in-house hardware hosting a file share, a cloud-based storage service, or a hosted web application. Although it is impossible to predict all future use case considerations, the following list provides a couple of important considerations:

1. A 30-minute WAV audio file with a 44.1kHz (16 bit depth) sampling rate takes ~92MB on disk (The Sustainable Heritage Network, 2015).
2. Transcription and diarization models are likely to continue to improve in accuracy over time. We encourage system developers to avoid designs that effectively limit the application of these models (e.g., compression of the data, reducing multiple microphone feeds into single track recordings). Also, if real time data analytics are required for a given use case, it is important to understand the real time factor of a system to determine the level of analysis that can be reasonably performed. The real time factor of an automated speech recognition system describes how many seconds are needed to process a single second of audio data. Thus, a real time factor of 1 means that the model can transcribe one second of audio in one second. (Srivastav et al., 2023).
3. In environments where security and data classification are crucial, networked storage systems introduce additional complexity with system security and approvals. Cloud-based systems impose different requirements than in-house hardware. Note: in human performance data contexts, cloud-based systems may trigger additional IRB requirements due to consent and data protection requirements.

Natural Language Processing

Several metrics characterize the structure of information exchange (e.g., word count, speech frequency measures, communication pathway analyses). These can lead to useful insights into *how* teams communicate, but not *what* is being shared. Natural language processing (NLP)-based text classification and dialogue act recognition allows for analysis of the content of messages shared between team members, allowing insight into aspects of team behavior (e.g., information sharing efficiency). NLP can support quantitative and qualitative investigation of decision making and behavioral processes at the individual or team level. Content-based analyses of communications or think-aloud protocols can provide useful insight into how speakers process information and conceptualize tasks, but decisions made regarding how data are collected and what elements are included can affect how easily such analyses are applied.

NLP requires accurate representations of message content for analysis. It is therefore important to ensure the communication modality used accurately captures the messages shared within the team. Text-based interfaces, such as chat, text messaging, and short-answer response formats, have the advantage of inherently generating an accurate transcript of team interactions. If audio communication is to be used, care must be taken to ensure good quality to facilitate later transcription and analysis. Poor audio quality can lead to transcripts with missing data due to inaudible or unintelligible speech. If possible, audio data should be collected with equipment that minimizes background noise and maintains consistent quality regardless of the actions of the speaker (e.g., moving around a space).

NLP analyses can be facilitated by the structure of the data. Some NLP analyses examine the timing of speech. Communication systems that incorporate timing into the data stream are extremely useful for later analysis. If using an open channel system, some sort of master timer to align to the data can be helpful. More resolution is generally preferred if the data context allows (e.g., time stamps for each sentence are better than time stamps for each conversational turn, and time stamps for the beginning and end of each sentence/utterance are better than time stamps only at the beginning). Similarly, data must support time alignment across multiple channels/sources. Events in one data stream (e.g., a simulator) should be relatable to events in another (e.g., a radio channel). If data streams cannot be synced directly to one another, other means such as event markers are useful.

NLP-driven communication analysis requires knowledge of not only what was said, but who said it. Sender identification is a key component of NLP in a team context. Possible means to facilitate speaker identification include separate communication channels for each speaker or standardized communication practice (e.g., starting each message with an address such as “Doc to Nurse. I need…”). A tagging procedure such as starting each recording with all participants identifying their role can also facilitate post-hoc assignment of speakers to each utterance.

Communication step-ons in which team members try to talk over one another can lead to messy transcripts that are difficult to construct and interpret. Alternatively, step-ons can be viewed as a potential outcome measure for analysis. If the situation allows control over communication systems, the investigator should think deliberately about selecting a system that allows step-ons to occur or how such events will be treated in the data stream. For instance, open channels may allow step-ons that are not easily transcribed. Systems that leverage multi-channel recordings allow the team to experience a step-on but still facilitate accurate transcription of all speakers’ utterances with overlapping timestamps.

Video Data Requirements

Capture of structured video data is critical for accurate and comprehensive analysis and is common in medical team training. This involves collecting quality video of the operational environment with appropriate context around the physical characteristics of the space, the tasks being executed, and metrics utilized to assess quality in performance. To ensure that the video data is suitable for driving analytics and computer vision processes, several best practices and considerations should be followed. First, using high-resolution cameras is essential to capture detailed movements and expressions, with a recommended resolution of at least 1080p (Aghajanzadeh et al., 2020). The frame rate should be sufficient to avoid motion blur, with 30 frames per second (fps) being a minimum standard, though higher frame rates like 60 fps can provide even more detail.

Next, it is important to record video under optimal lighting conditions when appropriate. Proper lighting not only enhances the visual quality of the footage but also enables more accurate tracking and analysis of movements, gestures, and expressions. A combination of natural and artificial light sources can be utilized, with careful positioning and diffusion to minimize shadows, glares, and low-contrast areas that can hinder analysis. Understanding that medical and military operations are often performed in a multitude of lighting conditions, it is important to continue video capture under these contexts to help improve current algorithms to account for non-ideal recordings.

Proper positioning and setup of recording equipment are also critical. Cameras should be placed at angles that capture the most relevant activities without obstruction, with careful consideration for the activities that require monitoring for analysis purposes. This requires decisions on placement of device, which can include egocentric (i.e., body worn camera for first-person perspective) or exocentric (i.e., external to the performers’ view with reference to the external environment). Exocentric placement can involve strategic locations for fixed cameras, while also being fixed to either a human or drone observer providing dynamic footage. For example, in medical training, cameras might be positioned to focus on hand movements, facial expressions, and the interaction between trainees and medical instruments.

Beyond proper setup within the training environment, adhering to relevant standards and guidelines is important to ensure that video data is collected securely, consistently and with analysis in mind. While specifics for the capture of video to drive analysis do not exist, some general IEEE and ISO (International Standards Organization) standards provide guidance. For instance, IEEE 802.1 addresses network capabilities for real-time video transmission, ensuring that high-quality video data can be reliably transferred and accessed (Hofmann, Nikolić & Ernst, 2019). In addition, ISO 23090-3:2021 addresses Versatile Video Coding, which is a method that can be leveraged to ensure efficient compression and transmission of high-resolution video data, enabling seamless capture and storage of high-quality footage for analysis (Hamidouche et al., 2022). Privacy and ethical considerations are also paramount when capturing video data for performance training and assessment. Obtaining informed consent from participants is essential, with clear communication about the purpose, scope, and intended use of the collected video data. Measures should be taken to protect the privacy and confidentiality of the data, such as anonymization, secure storage, and access controls, while ensuring compliance with relevant data protection regulations. This is critical when considering capture of video outside the oversight of IRBs and human-subjects research protections.

By leveraging AI, the military and organizations can streamline assessment processes, provide more personalized learning experiences, and gain deeper insights into student learning outcomes, ultimately enhancing the effectiveness and efficiency of video-based assessments in educational settings.

Detection and Tracking with Computer Vision

Preparing an effective computer vision algorithm for automated detection and tracking requires a multi-faceted approach. The foundation lies in the availability of high-quality and diverse training data, comprising annotated images or video frames that accurately label and segment the objects of interest (Everingham et al., 2010). This training data should encompass a wide range of scenarios, lighting conditions, and viewpoints to ensure the algorithm's robustness and generalization (Brock, De, Smith & Simonyan, 2021). Data augmentation techniques, such as rotation, flipping, scaling, and adding noise or occlusions, can be employed to artificially increase the diversity of the training data and enhance the algorithm's ability to handle variations in real-world scenarios (Shorten & Khoshgoftaar, 2019).

The selection and implementation of appropriate deep learning architectures and techniques are also crucial. Convolutional Neural Networks (CNNs) and object detection models like YOLO (You Only Look Once), Faster R-CNN, and Mask R-CNN have proven effective for object detection and tracking tasks (Li, Yang, Peng & Zhou, 2021). Additionally, techniques like optical flow analysis (Fortun et al., 2015) and temporal modeling (Kang et al., 2016) can be incorporated to improve the tracking capabilities of the algorithm, enabling it to maintain consistent object identities across frames and handle occlusions or rapid movements.

Beyond the training requirement, efficient data preprocessing and feature extraction pipelines are necessary to ensure the algorithm can process and analyze video data in real-time or near real-time. This may involve techniques like frame sampling, background subtraction (Bouwmans et al., 2019), and object tracking algorithms (Vatral et al., 2022) to reduce computational complexity and improve overall performance. Furthermore, continuous evaluation and fine-tuning of the algorithm are essential to ensure its accuracy and reliability in real-world deployment scenarios. This can involve techniques like cross-validation, performance metrics, and human evaluation to address biases or errors in the algorithm's predictions. Collaboration between domain experts, data scientists, and computer vision researchers is often necessary to develop and deploy successful automated detection and tracking solutions (Raghu et al., 2019).

Simulation Trace Data

Simulation trace data capture refers to the comprehensive collection of information generated by a simulation or game engine during its execution. This data encompasses multiple variables aligned to the state of the simulated environment, the sequence of user interactions, and performance metrics, providing a rich source for learning analytics. Some common examples of simulation trace data include:

1. State variables represent the state of the simulation at any given time, capturing attributes such as entity positions, speed/direction, health, and inventory statuses. State information also associates with weapons and systems that can be interacted with during scenario within a simulation environment.
2. Event data allows for the identification of start and stop times of specific events during a simulation-based training exercise, which include injects that drive a training and behavioral response. This information can

be invaluable for tracking and analyzing critical occurrences within the simulated scenario, with recent work focusing on the auto-detection of events for tagging and logging purposes (Goldberg et al., 2021).

3. Execution logs capture the sequence of actions taken by users, including navigation information, menu selections, and interactions with non-player characters (NPCs). These logs provide insights into the decision-making processes and behaviors exhibited by participants during the training exercise.
4. Performance metrics, such as task completion times, task accuracy, and other quantifiable variables, offer a means to measure and evaluate participant performance throughout the simulated exercise. These metrics can be crucial for assessing learning outcomes and identifying areas for improvement.

Collectively, these data streams generated by simulation trace data offer a comprehensive record of the training event, capturing the state of the simulation at various time points, user interactions with entities, and performance or criterion information. Logged at millisecond intervals, this data facilitates learning analytics by providing a rich source of actions, states, and assertions that can be analyzed to gain insights and inform instructional design. There are several current standards that drive the logging and sharing of trace data aligned to distributed simulation protocols. Those most commonly applied in the context of military training include Distributed Interactive Simulation (DIS; Hofer & Loper, 1995); High Level Architecture (HLA; Falcone, Garro, Anagnostou & Taylor, 2017), and Google Protocol Buffers (Currier, 2022). These standards provide an extensible approach for managing the capture and sharing of structured data sources around a defined schema. By leveraging simulation trace data, researchers and instructional designers can deeply understand the learning experiences within simulated environments, enabling them to optimize training scenarios, identify patterns, and ultimately enhance the effectiveness of simulation-based training programs.

Regardless of the standard that represents the simulation trace information, there needs to be a strategy to collect, contextualize the data through labeling, and then properly store the data and align it with other sources to drive analytics associated with learning and training outcomes. To ensure effective data capture, we recommend using existing data standards that offer a structured framework for organizing and representing information that can facilitate learner modeling. Standards such as the Experience API (xAPI) enable consistent tracking and communication of learning experiences across various platforms. The xAPI specification emphasizes interoperability, providing guidelines for diverse training environments to output performance data in a controlled manner for persistent storage and longitudinal modeling. This allows training systems and devices to operate within a broader ecosystem of training resources, supporting overall training progression toward readiness. Additionally, standards like SensorML and the OGC Sensor Observation Service (SOS) facilitate the capture of low-level and raw event data from various sensors and devices. Furthermore, time series data can be effectively captured using standards like OGC TimeseriesML, allowing for the storage and retrieval of time-stamped data for detailed analysis and visualization. By utilizing these standards, organizations can enhance interoperability, streamline data capture processes, and enable comprehensive analysis across different data sources and domains.

By leveraging machine learning algorithms, institutions can streamline assessment processes, provide more personalized learning experiences, and gain deeper insights into student learning outcomes, ultimately enhancing the effectiveness and efficiency of video-based assessments in training settings.

CONCLUSION

The demand for data availability within the DoD is crystal clear, yet today’s systems are lagging. Anticipation of data export and learning analytic needs is non-trivial but must be achieved to advance the art. Requirements for emerging training systems will be meaningfully informed only to the extent that human performance, learning analytics, and data needs are fully understood. This paper addresses a gap by presenting recommendations for a sample of specific data types critical for human performance analyses and by discussing the implications for system design and data export requirements. Our treatment of the topic is not exhaustive. We presented in-depth discussion related to audio, video, and simulation trace data to equip program managers and system developers with a perspective on design considerations that make-or-break future analytic capabilities. These are just a few types of relevant data to be considered. As always, more work is needed.

We urge designers of future systems to adopt these recommendations to maximize data availability in support of readiness and learning analytics. Pursuit of this goal requires careful consideration of data repository needs, an in-depth understanding of the criticality of metadata, and educated approaches to system design to avoid negative impacts to data quality. These considerations have tangible impact on the DoD's ability to perform advanced analytics and adapt to ever-changing operational conditions. As the warfighting context evolves, system designs that suffer from these known issues effectively limit access to data and contribute to the ways the DoD can be outpaced. Alternatively, proactive solutions built on an understanding of data analytics and a vision toward future capabilities will answer the DoD's call for data availability and will be key enablers in maximizing data leverage for decisional advantage.

ACKNOWLEDGEMENTS

The views expressed in this article reflect the results of research conducted by the authors and do not necessarily reflect the official policy or position of the Department of the Air Force, Department of Defense, the Department of the Army, the Department of the Navy, nor the United States Government. DISTRIBUTION A: Approved for public release. Distribution Unlimited. AFRL Case Number: AFRL-2024-3364.

REFERENCES

Aghajanzadeh, S., Naidu, R., Chen, S. H., Tung, C., Goel, A., Lu, Y. H., & Thiruvathukal, G. K. (2020, October). Camera placement meeting restrictions of computer vision. In 2020 IEEE International Conference on Image Processing (ICIP) (pp. 3254-3258). IEEE.

Bell, H. H. (1999). The effectiveness of distributed mission training. *Communications of the ACM*, *42*(9), 72-78.

Bouwmans, T., Javed, S., Sultana, M., & Jung, S. K. (2019). Deep neural network concepts for background subtraction: A systematic review and comparative evaluation. Neural Networks, 117, 8-66.

Brock, A., De, S., Smith, S. L., & Simonyan, K. (2021, July). High-performance large-scale image recognition without normalization. In International Conference on Machine Learning (pp. 1059-1071). PMLR.

Currier, C. (2022). Protocol buffers. In Mobile Forensics–The File Format Handbook: Common File Formats and File Systems Used in Mobile Devices (pp. 223-260). Cham: Springer International Publishing.

Defense Acquisition University (2013). DoD Open Systems Architecture Contract Guidebook for Program Managers: a tool for effective competition. Ft. Belvoir, VA.

Defense Health Agency. (2018). Defense Health Agency Procedural Instruction 6025.06: Standardized Templates for Primary Care Clinical Encounter Documentation (No. 6025.06). Retrieved from https://www.health.mil/-/media/Files/MHS/Policy-Files/Signed-DHAPI-602506-Standardized-Templates-for-Primary-Care-Clinical-E.ashx

Department of Defense (2023). Data, Analytics, and Artificial Intelligence Adoption Strategy: Accelerating Decision Advantage.

den Uijl, S., de Vries, H. J., & Bayramoglu, D. (2013). The Rise of MP3 as the Market Standard: How Compressed Audio Files Became the Dominant Music Format. International Journal of IT Standards and Standardization Research (IJITSR), 11(1), 1-26. http://doi.org.wrs.idm.oclc.org/10.4018/jitsr.2013010101

Everingham, M., Van Gool, L., Williams, C. K., Winn, J., & Zisserman, A. (2010). The pascal visual object classes (voc) challenge. International journal of computer vision, 88, 303-338.

Falcone, A., Garro, A., Anagnostou, A., & Taylor, S. J. (2017, December). An introduction to developing federations with the High Level Architecture (HLA). In 2017 Winter Simulation Conference (WSC) (pp. 617-631). IEEE.

Fortun, D., Bouthemy, P., & Kervrann, C. (2015). Optical flow modeling and computation: A survey. Computer Vision and Image Understanding, 134, 1-21.

Garcia, D. E., Mertens, A., Hernandez, J., Li, M., & Mann, S. (2020, October). HDR (high dynamic range) audio wearable and its performance visualization. In *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (pp. 3333-3339).

Goldberg, B., Owens, K., Gupton, K., Hellman, K., Robson, R., Blake-Plock, S., & Hoffman, M. (2021). Forging competency and proficiency through the synthetic training environment with an experiential learning for readiness strategy. In Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC), Orlando, FL.

Goodell, J., & Kolodner, J. (Eds.). (2022). *Learning engineering toolkit: Evidence-based practices from the learning sciences, instructional design, and beyond*. Taylor & Francis.

Gorman, J. C., Grimm, D.A., Stevens, R.H., Galloway, T., Willemsen-Dunlap, A.M., & Halpin, D.J. (2020). Measuring real-time team cognition during team training. *Human Factors, 62*, 825-860.

Gorman, J. C & Wiltshire, T.J. (2022). A typology for the application of team coordination dynamics across increasing levels of dynamic complexity. *Human Factors*, *66*(1), 5-16.

Greenberg, J., White, H. C., Carrier, S., & Scherle, R. (2009). A Metadata Best Practice for a Scientific Data Repository. *Journal of Library Metadata*, *9*(3-4), 194-212.

Guertin, N., & Hurt, T. (2013). DoD Open Systems Architecture Contract Guidebook for Program Managers. *Defense AT&L, September-October*, *30*, 32.

Hamidouche, W., Biatek, T., Abdoli, M., François, E., Pescador, F., Radosavljević, M., ... & Raulet, M. (2022). Versatile video coding standard: A review from coding tools to consumers deployment. IEEE Consumer Electronics Magazine, 11(5), 10-24.

Harada, N., Moriya, T., & Kamamoto, Y. (2007). Mpeg-4 als: performance, applications, and related standardization activities. *NTT Technical Review*. https://www. ntt-review. jp/archive/ntttechnical. php.

Harrison, J. L., Gorman, J. C., Reitman, J. G., & D'Mello, S. (2023). Toward the Validation of a Novel Measure of Individuals' Influence During Team Collaborations. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 67*(1), 290-295. https://doi.org/10.1177/21695067231192643

Hernandez, M., Goldberg, B., Robson, R., Owens, K., Blake-Plock, S., Welch, T., & Ray, F. (2022). Enhancing the Total Learning Architecture for Experiential Learning. In *Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC)*.

Hofer, R. C., & Loper, M. L. (1995). DIS today [Distributed interactive simulation]. Proceedings of the IEEE, 83(8), 1124-1137.

Hofmann, R., Nikolić, B., & Ernst, R. (2019). Challenges and Limitations of IEEE 802.1 CB-2017. IEEE Embedded Systems Letters, 12(4), 105-108.

IASA Technical Committee, *Guidelines on the Production and Preservation of Digital Audio Objects*, ed. by Kevin Bradley. Second edition 2009. (= Standards, Recommended Practices and Strategies, IASA-TC 04). www.iasa-web.org/tc04/audio-preservation

IEEE Standard Adoption of Moving Picture, Audio and Data Coding by Artificial Intelligence (MPAI) Technical Specification Context-based Audio Enhanced (CAE) Version 1.4. (2023). IEEE Std 3302-2022, 1–94. https://doi.org/10.1109/IEEESTD.2023.10112597

IEEE Standard for Advanced Audio Coding. (2023). IEEE Std 1857.2-2023 (Revision of IEEE Std 1857.2-2013), 1–349. https://doi.org/10.1109/IEEESTD.2023.10258051

IEEE Standard for Distributed Interactive Simulation–Application Protocols. (2012). IEEE Std 1278.1-2012 (Revision of IEEE Std 1278.1-1995), 1–747. https://doi.org/10.1109/IEEESTD.2012.6387564

IEEE Standard for Second Generation Audio Coding. (2020). IEEE Std 1857.8-2020, 1–470. https://doi.org/10.1109/IEEESTD.2020.9271961

Kang, K., Ouyang, W., Li, H., & Wang, X. (2016). Object detection from video tubelets with convolutional neural networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 817-825).

Krumm, A., Means, B., & Bienkowski, M. (2018). *Learning analytics goes to school: A collaborative approach to improving education*. Routledge.

Li, Z., Liu, F., Yang, W., Peng, S., & Zhou, J. (2021). A survey of convolutional neural networks: analysis, applications, and prospects. IEEE transactions on neural networks and learning systems, 33(12), 6999-7019.

Marler, T., Lewis, M. W., Toukan, M., Haberman, R., Kochhar, A. K., Downing, B., ... & RAND CORP ARLINGTON VA. (2021). *Supporting Joint Warfighter Readiness: Opportunities and Incentives for Interservice and Intraservice Coordination with Training-Simulator Acquisition and Use*. RAND.

Martin, E.G.; Law, J.; Ran, W.; Helbig, N.; Birkhead, G.S. (2017). Evaluating the quality and usability of open data for public health research: A systematic review of data offerings on 3 open data platforms. *J. Public Health Manage. Pract. 23*, 5-13.

Naval Air Warfare Center Training Systems Division (NAWCTSD) & Cole Engineering. (2021). Military Health System Interoperable Networking for Training, Readiness, and Education in Medicine (INTREMED) Analysis of Standards.

Naval Air Warfare Center Training Systems Division (NAWCTSD) & Katmai Corporation (2023) Military Health System Interoperable Networking for Training, Readiness, and Education in Medicine (INTREMED) Data Strategy and Architecture Report.

Naval Air Warfare Center Training Systems Division and Cole Engineering, "Analysis of Standards- Part A Summary of Findings and Recommendations," Sep. 2021.

Ng, P. S., & Sanches, I. (2004). The influence of audio compression on speech recognition systems. In 9th Conference Speech and Computer.

Park, T. J., Kanda, N., Dimitriadis, D., Han, K. J., Watanabe, S., & Narayanan, S. (2022). A review of speaker diarization: Recent advances with deep learning. *Computer Speech & Language*, *72*, 101317.

Perazzo, J., Rodriguez, M., Currie, J., Salata, R. & Webel, A. R. (2019). Creation of Data Repositories to Advance Nursing Science. *Western Journal of Nursing Research*, 4(1), 78-95.

Raghu, M., Blumer, K., Corrado, G., Kleinberg, J., Obermeyer, Z., & Mullainathan, S. (2019). The algorithmic automation problem: Prediction, triage, and human effort. arXiv preprint arXiv:1903.12220.

Reynolds, D. A., & Torres-Carrasquillo, P. (2005). Approaches and applications of audio diarization. Proceedings. (ICASSP '05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005., 5, v/953-v/956 Vol. 5. https://doi.org/10.1109/ICASSP.2005.1416463

Schreiber, B. T. (2013). Transforming Training: A Perspective on the Need for and Payoffs from Common Standards. *Military Psychology*, *25*(3), 294-307.

Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. Journal of big data, 6(1), 1-48.

Srivastav, V., Majumdar, S., Koluguri, N., Moumen, A., Gandhi, S., Hugging Face Team, Nvidia NeMo Team, & SpeechBrain Team. (2023). *Open Automatic Speech Recognition Leaderboard*. Hugging Face. https://huggingface.co/spaces/hf-audio/open_asr_leaderboard

The Sustainable Heritage Network. (2015, March 20). *Estimating Storage Requirements for Digital Audio*. https://sustainableheritagenetwork.org/digital-heritage/estimating-storage-requirements-digital-audio

Trisovic, A.; Mika, K.; Boyd, C.; Feger, S.; Crosas, M. (2021). Repository Approaches to Improving the Quality of Shared Data and Code. *Data*, *6*(15).

Vatral, C., Biswas, G., Cohn, C., Davalos, E., & Mohammed, N. (2022). Using the DiCoT framework for integrated multimodal analysis in mixed-reality training environments. *Frontiers in artificial intelligence*, *5*, 941825.

Walcutt, J.J. & Schatz, S. (Eds.) (2019). Modernizing Learning: Building the Future Learning Ecosystem. Washington, DC: Government Publishing Office.

Wydler, V. (2014). Gaining Leverage Over Vendor Lock to Improve Acquisition Performance and Cost Efficiencies. 1–20. https://www.dair.nps.edu/bitstream/123456789/1081/1/SYM-AM-14-033.pdf