A Data-Driven Organization

Developing data and information competency

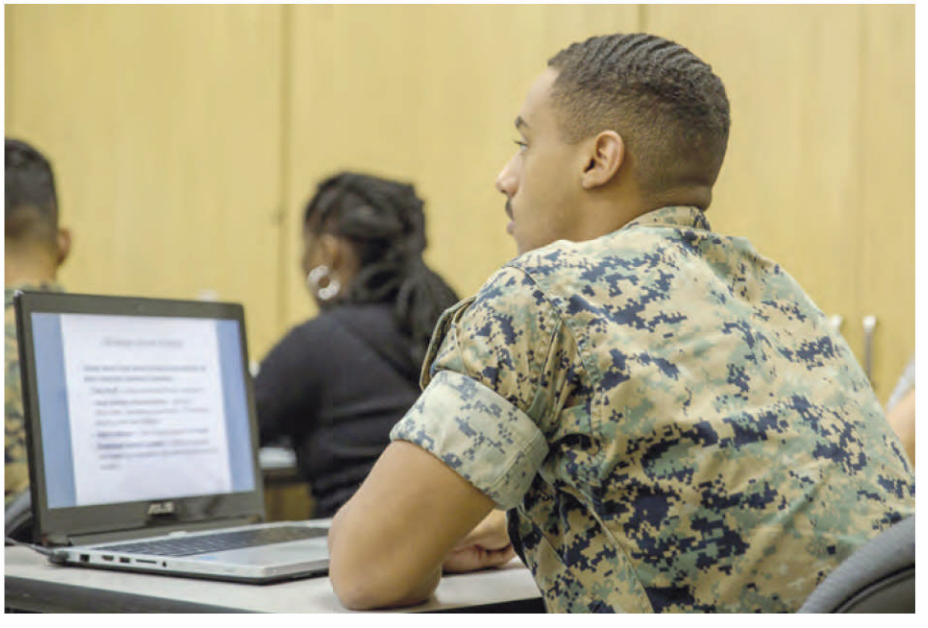
by Maj Scott A. Humr

The Marine Corps is a data driven organization and therefore requires a greater number of Marines to possess the knowledge, skill, and abilities of data and information competencies to remain effective on the battlefields of today and the future.[1] However, minimal training is devoted to increasing the proficiency of Marines across all ranks to further develop the knowledge, skills, and abilities required to be effective in an environment characterized by over-the-size-limit email inboxes, duplicative systems of record, unstructured share drives, and disorderly data management practices. Therefore, the Marine Corps must inculcate data and information competency instruction across all modalities of training and education. Our bid for success is to take a systems approach to develop these competencies in our Marines which will have second and third order effects that yield increases in data and information accuracy and, ultimately, a more lethal Marine Corps.

>Maj Humr is a Ground Supply Officer. He currently serves as the Graduate Education Manager at MMOA-3.

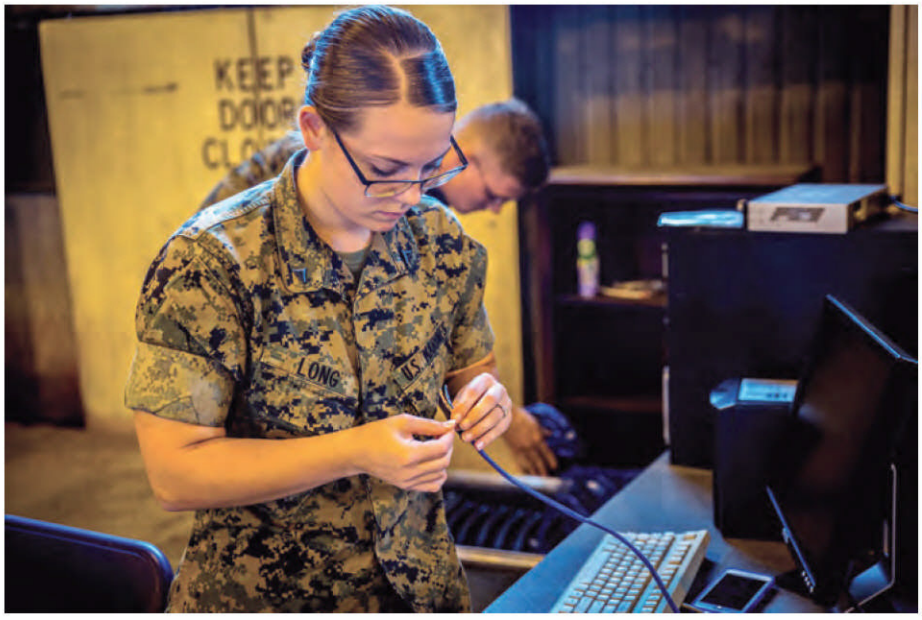
The Current Environment

Our most important processes require data, information, or knowledge to make informed decisions and take appropriate actions for accomplishing any mission. From planning processes at the tactical edge requiring target intelligence to strategic resource decisions on long-term budget line items, the importance of accurate information cannot be emphasized enough. For instance, a recent RAND report for the intelligence community asserted that, "[e]xploiting the rapidly growing sources of data available for collection and analysis is one of the greatest professional crises facing today's intelligence leaders."[2] Parallels from the RAND report are easily drawn to any community that desires to discover key insights and relies on data to make important decisions. In fact, the Commandant's responsibility to man, train, and equip the force for world-wide deployment or other taskings could not be performed successfully if he did not have an accurate feedback mechanism for measuring the efficacy of his direction and guidance toward producing the capabilities he is tasked to provide. For this reason, the Marine Corps enterprise is constantly producing, sorting, sending, searching, and storing an incomprehensible amount of data and information, which is continually refined at higher levels in the organization to provide this feedback. Therefore, we must imbue in our Marines the importance of data and information by providing the necessary training to make them and our organizations successful.



The Marine Corps is a data driven organization. *(Photo by LCpl Mason Roy.)*

The routine actions and behaviors conducted by Marines on a daily basis are producing data and information that feed higher levels of reporting. For example, lance corporals becoming proficient in their training and readiness standards to the sergeant entering supply transactions into Global Combat Support System-Marine Corps are actions that influence the readiness status of personnel and equipment captured in various systems of record across the



Information is gathered from multiple sources, saved, and shared with other users. *(Photo by LCpl Andy Martinez.)*

enterprise. Most importantly, the cumulative effects of all actions stored in these systems provide an assessment of our true capabilities as an organization available to a geographic combatant commander.

Regrettably, many of our organizational data environments are characterized by a great deal of disorganization, superfluous files, amorphous information, and a variety of bespoken applications, which often replicate the inefficient paper processes. However, this is what we should expect given most Marines receive practically no training on how to use many of the tools they are provided or the paltry online records management course mandated once a year. Marines kludge information from multiple data sources, saving and sharing the data across the enterprise. Versions of information proliferate and spawn subsequent variations with many iterations quickly becoming out-of-date. There is no wonder as to why we find ourselves in this most unenviable position.

As Marines, we take many things for granted when it comes to our expectations for what constitutes a "basically trained Marine." We expect a Marine arriving to his first fleet assignment to understand rudimentary military concepts such as history of the Marine Corps, chain of command, close order drill, weapons safety, and customs and courtesies; we also presume Marines will know the basics of his MOS. However, many Marines within administration and logistics functions will often fall in on numerous spreadsheets, out-of-date slide presentations, and government email accounts. An often-misinformed notion is that today's Marines have a natural penchant for these tools given they are stereotyped as "digital natives." While this may be true, it does not necessarily mean that they know how to use even the elementary capabilities of these tools provided to them, let alone the more advanced capabilities required to effectively accomplish many of the routine tasks required of them. We have all seen the Marine reconciling two spreadsheets side-by-side on a screen or printed copy using his finger to confirm changes or spot anomalies. Worse, Marines will use a separate calculator when adding a column of numbers in a spreadsheet, unaware that this and many other functions are inherently available to them. Recall rosters are distributed to a variety of individuals over email for one unfortunate Marine to cut and paste into the "master" spreadsheet once the various versions are returned. These examples, amongst many others, are the product of an indifferent and ineffective leadership. Such disregard results in countless man-hours wasted by our Marines; man-hours which could be better spent on more meaningful work and activities that are both rewarding and impactful for our organizations.

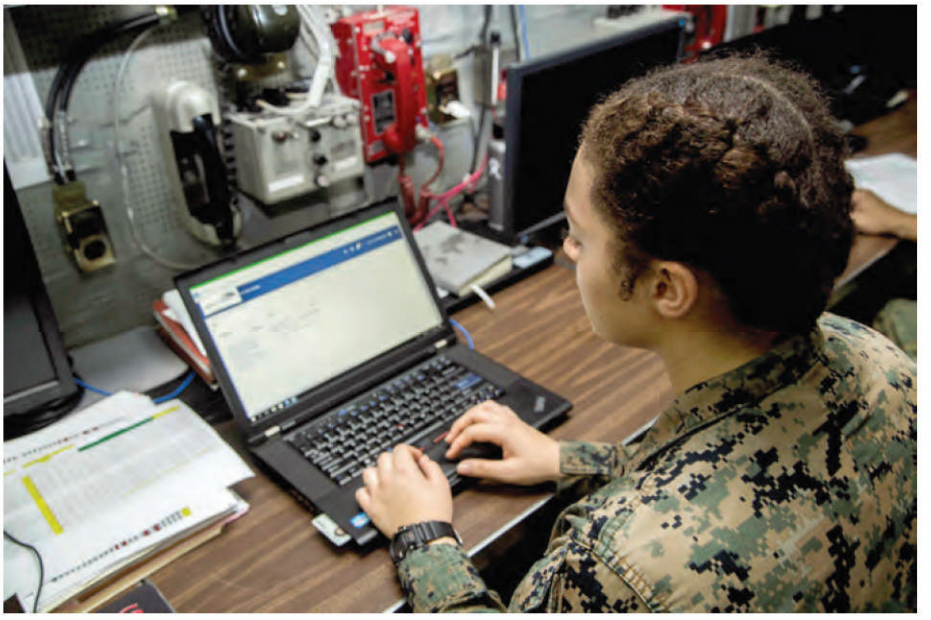
A Better Way

Marines must be able to adapt to efficiently manipulate and utilize programs such as Excel or Access. In fact, the Marine Corps is currently in possession of a suite of software applications that make data analysis much easier. Yet, Marines are still handicapped by their own ignorance of the availability of these applications combined with a lack of training. Unfortunately, it is too common to see Marines doing things the hard way when it comes to data collection, manipulation, cleansing, and analysis. This must change.

The velocity, volume, variety, and veracity (4Vs) of data within organizations are constantly increasing,[3] which will require an educated workforce that understands the fundamentals of data.[4] If the Marine Corps truly values data as a key driver in understanding combat readiness and performance, it must inculcate a data science mindset across its force. Furthermore, data and information are not the exclusive purview of any particular individual—all Marines are responsible.

To better equip our Marines, the Marine Corps needs to create a data-focused curriculum that will become a periodic training requirement or result in a bona fide certification. The creation of a curriculum that focuses our efforts on the importance of data is not only more empowering, it is needed to help drive innovations and decision making in the future. The proposed curriculum will be composed of, but not limited to, topics such as:

- Introduction to the basics of data science.
- Database fundamentals.
- Basic statistics and analysis.
- The importance of data and how it helps drive decision making.
- Data provenance and information management.



Data information must be updated and readily accessible in order to prepare reports or for easy access by the commander. *(Photo by LCpl A.J. Van Fredenberg.)*

- Data/information/knowledge hierarchy.
- Data visualization techniques.
- Dashboards.
- Data ontologies and taxonomies.
- Extract/transfer/load tools and techniques.
- Basic Excel, Access, Power BI, Tableau, R.
- Data cleansing and manipulation techniques and best practices.
- Logistics focused module on manpower or logistics automated information systems and their data structures.
- Data munging/wrangling/collection.

Many of these topics are already available through SkillSoft courseware on MarineNet and through free online learning platforms such as Coursea[5] and Udacity.[6] The Marine Corps may also potentially leverage the U.S. Army's Operations Research and Systems Analysis program, Naval Postgraduate School Distance Learning, and the Air Force Institute of Technology courses as additional venues for further training. Additional MOS-specific tracks could be created to further define the necessary skills for particular environments or billets.

Conclusion

Creating data literacy across the enterprise is key to ensuring the Marine Corps is postured to understand the importance of their data and how it provides value to the entire organization. Proper organization of our critical information for fast retrieval and data literacy is fundamental for improving our problem-solving abilities and solutions. For these reasons, educating the workforce on the skills and tools they need to curate and understand data will make large strides toward improving data-driven decisions across the Marine Corps. New technologies require large volumes of curated data to train and test appropriately.[7] Therefore, if the Marines Corps desires to leverage technologies, such as artificial intelligence and machine learning, it will need to both understand the data and properly organize it for consumption and training by the algorithms that fundamentally require vast amounts of data. Looking ahead, one may envision an additional MOS for Marines who have earned these credentials in their MOS for being subject matter experts on their field's particular data systems. For instance, the logistics community could create a 04XX AMOS similar to the 0477 or the 8055 Information Management Officer additional MOS as well. Particular billet MOSs might one day be coded with an additional skills designator for a Marine with particular data science skill set.

These skills and technologies are already within our reach and available to all Marines. Leaders need to prioritize these skill sets if the Marine Corps is to be successful in the future. Leaving these skills to hope and chance is not a course of action for success. Rather, they are outdated blueprints for status quo indifference that will continue to enfeeble our efforts toward increased lethality. To thrive as an institution in the current and future environment, we must prepare for it by training our Marines accordingly. Anything less is gross negligence and a prescription for defeat.

Notes

1. Andrew Churchill, "Data Analytics Will Fuel the Future of Military Readiness," *Defense Systems*, (Online: June 2018), available at https://defensesystems.com/articles.

2. Bradley Knopp, Sina Beaghley, Aaron Frank, Rebeca Orrie, Michael Watson, "Defining the Roles, Responsibilities, and Functions for Data Science Within the Defense Intelligence Agency," RAND Corporation, (Online: 2016), available at https://www.rand.org.

3. Cortney Weinbaum and John N.T. Shanahan, "Intelligence in a Data-Driven Age," *Joint Forces Quarterly*, (Washington, DC: National Defense University Press, July 2018), available at https://ndupress.ndu.edu.

4. Jonathan Corneliessen, "The Democratization of Data Science," *Harvard Business Review*, (Brighton MA: Harvard Business Publishing, July 2018), available at https://hbr.org.

5. Coursera, "Data Science Specialization," *Coursera*, (Online), available at https://www.coursera.org.

6. Udacity, Data Science Nano Degree, *Udacity*, (Online), available at https://www.udacity.com.

7. "Is There Too Much Data for Machine Learning?" *IEEE*, (Online: May 2018), available at https://transmitter.ieee.org.