

Managing T&E Data to Encourage Reuse

Andrew Flack

Institute for Defense Analyses, Alexandria, VA

Reusing Test and Evaluation (T&E) data sets multiple times at different points throughout a program's life cycle is one way to realize their full value. Data management plays an important role in enabling—and even encouraging—this practice. Although Department-level policy on data management is supportive of reuse and consistent with best practices from industry and academia, the documents that shape the day-to-day activities of T&E practitioners are much less so. As a result, reuse of T&E data sets does not occur on a consistent basis or in a formalized way. To fill this apparent gap, this article expands upon four best practices—addressed in different ways in Service-specific T&E policies—that can increase the reuse of T&E data sets.

Value of T&E Data

Test and Evaluation (T&E) data sets are valuable resources. When collected in operationally realistic environments, they are rich sources of knowledge that help answer questions about the effectiveness, suitability, and survivability of military systems. Since the 2010 "Guidance on the Use of Design of Experiments (DOE)" from the Director, Operational Test and Evaluation (DOT&E), rigorous test design has increased the knowledge content of T&E data sets by strategically placing test points in important parts of the operational envelope (DOT&E 2010). In the analysis phase, new statistical methods have improved extraction of that knowledge, helping analysts do more with the data they have. Clearly, the T&E community has made progress in this area, but we still do not realize the full value of T&E data sets—we should reuse them multiple times throughout a program's life cycle, extracting value with each use.

Data is a unique resource in that its value does not diminish each time it is processed. As such, data's full value is much greater than the value extracted from its first use (Mayer-Schönberger and Cukier 2013). In the T&E community, we tend to focus on this first use—to perform an evaluation ahead of an upcoming milestone decision, for example. But if we set the data set aside when the analysis is complete, we leave the residual value untapped.

The T&E community should strive to reuse T&E data sets on a more consistent and formalized basis. In



Andrew Flack

doing so, not only will we realize value from those data sets many times over, but we may also:

- Shift system evaluations left in a program's life cycle by making use of early program test data. If we can reuse data from developmental tests (DT), we might be able to do less operational testing (OT), for example.
- Reduce or eliminate redundancies associated with collection of data that may exist elsewhere.
- Do better science—increase the transparency and credibility of analyses, encourage innovation, and perform richer analyses.

Data management plays a key role in enabling, and even encouraging, reuse of T&E data sets.

The Role of Data Management

First, it is important to clarify the term "reuse." In their 2017 article, "On the Reuse of Scientific Data," Pasquetto, et al. (2017) distinguish "use" and "reuse" by noting that reuse implies usage of the data set by someone other than the originator. They point out that reuse is typically discussed either in the context of reproducibility, where "data from a prior study are reanalyzed to validate, verify, or confirm previous research," or integration, where data sets are reused in combination with other data to "make comparisons, build new models, or explore new questions altogether." In this article, we focus on the latter definition of reuse: the

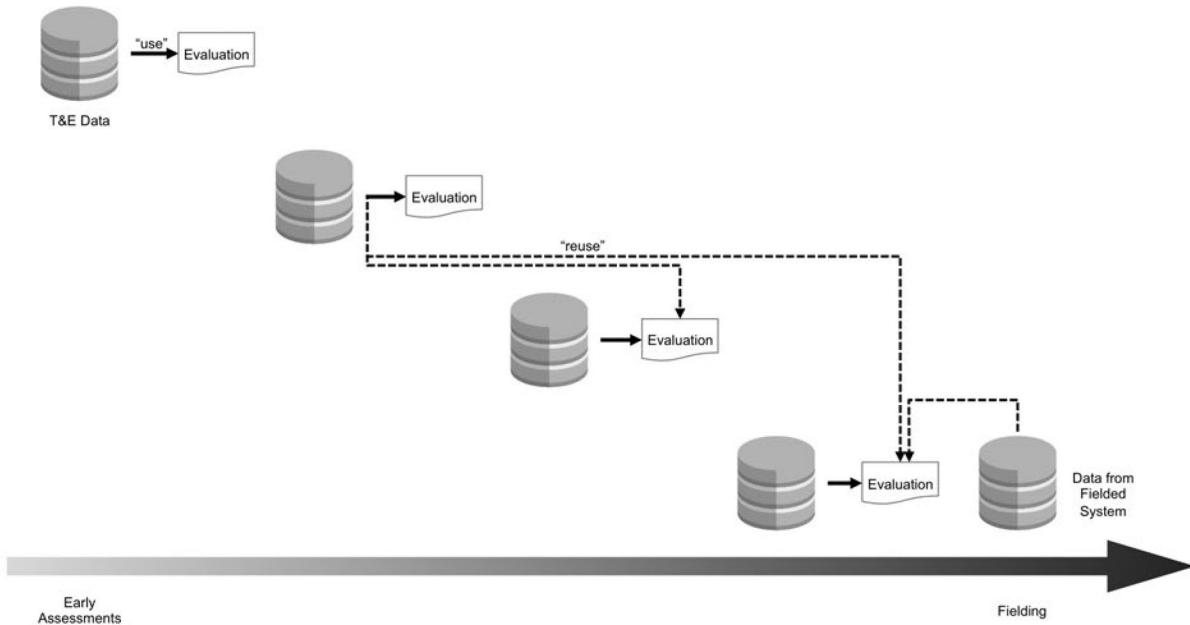


Figure 1: Data Reuse Schematic

integration of multiple existing T&E data sets, or the integration of existing T&E data sets with new T&E observations. Figure 1 provides a notional schematic for this type of reuse.

Given this definition, a scenario (albeit overly simplified) may help demonstrate the role of data management in enabling reuse of T&E data sets:

An analyst is preparing to perform an assessment of a radar system following a recent operational test. One aspect of the assessment is an evaluation of the radar's ability to classify different munition types. During the test, light and heavy artillery, light and heavy mortars, and light rockets were fired. Because of test constraints, heavy rocket trials were scratched.

To fill this gap in the assessment, the analyst must have the capability to perform the following steps:

1. Search in a dedicated T&E database for OT or DT data sets where similar platforms were tested against heavy rockets.
2. If such a data set is found, inspect the metadata, or “data about data,” detailing the data set’s context, assumptions, and limitations. Where and when was the test conducted? What were the environmental conditions? What were the system configuration and mission type?
3. Given this information, make a determination about whether the data set is acceptable for use in the evaluation.
4. Access and use the data set.

Department-level policies and guidance on data management support the scenario above and are consistent with industry and academia best practices for enabling discovery and reuse. However, Department of Defense (DoD) Instruction (DoDI) 5000.02, “Operation of the Defense Acquisition System,” and Defense Acquisition University’s Defense Acquisition Guidebook (DAG), two documents that shape the day-to-day work of T&E practitioners, are much less supportive of reuse. To fill this apparent gap, we expand upon four best practices—addressed in different ways in Service-specific T&E policies—that can increase the reuse of T&E data sets.

DoD Policy and Guidance on Data Management

The DoD has issued several strategic directives and guidance documents that establish and evolve policies for the management, use, and sharing of DoD data:

- DoD Data Administration, September 1991 (DoD 1991)
- DoD Net-Centric Data Strategy, May 2003 (DoD CIO 2003)
- Data Sharing in a Net-Centric Department of Defense, December 2004 (DoD 2004)
- Guidance for Implementing Net-Centric Data Sharing, April 2006 (DoD 2006)
- DoD Net-Centric Services Strategy (NCSS), May 2007 (DoD CIO 2007)

- Sharing Data, Information, and Information Technology (IT) Services in the Department of Defense, August 2013 (DoD 2013)
- Implementing the Sharing of Data, Information, and Information Technology (IT) Services in the Department of Defense, August 2015 (DoD 2015)

The Net-Centric Data Strategy (NCDS), published by the DoD Chief Information Officer (CIO) in 2003, is the most notable, as it marked a shift in the Department's data management approach away from "data administration." Standardizing and controlling data elements, definitions, and interfaces across the Department was too cumbersome; the Department is too large and the data too varied for this approach to be sustainable.

Instead, the NCDS proposed two overarching objectives that established a new paradigm for data management that supports reuse. First, increase the amount of data available to communities of interest and the enterprise by focusing on the visibility and accessibility of data, rather than standardization. Second, ensure that data are usable by both anticipated and unanticipated users through extensive use of metadata.

The NCDS was codified in 2004 by DoD Directive (DoDD) 8320.02, "Data Sharing in a Net-Centric Department of Defense," with policies, responsibilities, and implementation procedures established by DoDI 8320.07 in 2015. DoDI 8320.07 ensures that data, information, and IT services are "visible, accessible, understandable, trustworthy, and interoperable," throughout their life cycles for all authorized users. As in the NCDS itself, the procedures describe extensive use of metadata to support these goals: make data "visible" through the use of tags and other so-called "discovery metadata," "accessible" by allowing and controlling access through security-related metadata, "understandable" through structural or semantic metadata, and "trusted" by including pedigree metadata.

Approaches from Industry and Academia

Although the language is different, the concepts described in the 2003 NCDS—that data be visible, accessible, understandable, trustworthy, interoperable, and usable by both anticipated and unanticipated users—are consistent with "FAIR," one of the most commonly referenced data management frameworks from industry and academia.

"FAIR" is a set of principles that aims to improve the infrastructure supporting the discovery and reuse of scholarly data. Published in a 2016 article in *Scientific Data* by Wilkinson et al., the FAIR Guiding Principles

—Findability, Accessibility, Interoperability, and Reusability—support deriving the maximum long-term value from research data (Wilkinson 2016). The authors loosely define the individual principles in order to increase flexibility and make it easy for data producers, publishers, and stewards to work toward "FAIRness."

Data are "Findable" when the data set and supplementary materials have rich discovery metadata and a unique persistent identifier and are indexed in a searchable resource. The Google developers' guide for discovery of data sets offers a good example of this FAIR Guiding Principle in application (Google n.d.). The company encourages data producers to provide structured information based on the Schema.org Data set markup (Data set n.d.) or the World Wide Web Consortium's (W3C) Data Catalog Vocabulary (DCAT) format (Government Linked Data Working Group 2014). These metadata standards are widely accepted and used.

Discovering the existence of a data set is not enough to be able to reuse it; data and metadata must also be "Accessible." Accessible data are retrievable without specialized tools or communication methods and have metadata that remain available even when the data are sensitive, have special access considerations, or are no longer publicly available. Keeping metadata accessible acknowledges the existence of a data set and allows the user to locate the data, perhaps by finding a point of contact or the name of a relevant organization.

Finally, data are "Interoperable" when both the data and metadata follow formal, widely applicable standards for knowledge representations (e.g., HTML, JSON, XML, or CSV), and "Reusable" when they include all information necessary to understand how they can be used in a new context, such as rights (intellectual property or other) and licensing.

T&E-specific Policy and Guidance

Although Department-level policy on data management is supportive of reuse, the documents that shape the day-to-day work of T&E practitioners are much less so. Program managers (PMs) execute programs in accordance with DoDI 5000.02, "Operation of the Defense Acquisition System," and the non-mandatory guidance described in the Defense Acquisition University's Defense Acquisition Guidebook (DAG) (DoD 2015, DAU 2017). These documents do not reflect either the policies under DoDD 8320.02 or the guidance under DoDI 8320.07 as strongly as they should; the NCDS considers the core of the net-centric environment to be any data that enable effective decisions, and T&E data sets certainly meet this criterion.

Specifically, language from DoDI 8320.07 to make data “visible, accessible, understandable, trusted, and interoperable” is lacking in the DAG; the “Test and Evaluation” chapter should certainly include this language. Similarly, DoDI 5000.02 provides little guidance with respect to the collection, management, and storage of T&E data, except to point out the use of a Test and Evaluation Master Plan (TEMP) as the primary planning and management tool.

Fortunately, Service-specific T&E policies offer guidance that fills this apparent gap, aligning the management of T&E data with guidance from DoDI 8320.07 and DoDD 8320.02. All Service-specific T&E policies:

- Recommend a common T&E database to share program test data across T&E stakeholders. (Only Air Force Instruction (AFI) 99-103, Capabilities Based Test and Evaluation, names it a “Common T&E Database” and emphasizes access to program stakeholders with a need-to-know; the other Service policies address the subject more obliquely.)
- Discuss planning for data needs, with Army policy and guidance (Army Regulation (AR) 73-1, Test and Evaluation Policy, and Department of the Army Pamphlet (DA Pam) 73-1, Test and Evaluation in Support of Systems Acquisition) providing the most robust discussion on curation, authentication, and disposition.
- Support the use of data and findings from DT to supplement OT data in system evaluation, with some also including data from contractors, deployed assets used in real-world operations, and other sources.
- Emphasize the importance of data pedigree, including archival of other documentation alongside T&E data that may be of value in understanding the complete historical picture of a program.

Recommendations for the T&E Community

Among Service-specific policies, there is variation in the depth of discussion of specific data management topics, the extent to which they reflect the policies and practices outlined in 8320.07, and the extent to which, as written, they would support reuse of T&E data. However, taken together, they fill in the gaps in DoDI 5000.02 and the Defense Acquisition Guidebook. By expanding on the common themes across Service policies, we can make four high-level recommendations that will encourage increased reuse of T&E data sets: (1) Establish a program-level T&E database as early as possible, (2) Focus on rich metadata, (3) Strive to store raw

or minimally processed data, and (4) Consider your data’s life cycle.

Establish a Program-level T&E Database, and Plan for it Early

Program managers should establish a common T&E database and manage it throughout the program’s life cycle. The database should be visible and accessible to all stakeholders with a need-to-know, to include all members of the T&E Working-level Integrated Product Team (WIPT). Managing the database includes curation, managing access, and developing strategies for disposition (or archival) at the program’s conclusion. Data ownership at the program level allows flexibility to implement any data management system that works best for the specific program.

Data management to encourage reuse should be considered early and continuously throughout the entire life cycle of a program. In fact, Service policies direct the program manager to create a common T&E database as early as practical (see, for example, SEC-NAV-M 5000.2, Department of the Army Pamphlet 73-1, and Air Force Instruction 99-103). It is critical to establish the common T&E database early in order to capture the earliest available test data, such as contractor data (where possible) and early DT data.

Planning for a program-level T&E database should occur as part of the preparation of the Life Cycle Sustainability Plan (LCSP) (or the earliest practical acquisition document for those programs not following the standard Joint Capabilities Integration and Development System (JCIDS) process) and the TEMP. The LCSP is required starting at Milestone A and describes the technical requirements and management activities involved in sustaining a system. Program managers should plan and resource data management as part of the LCSP. Using this early planning document not only ensures that data management will be considered early on, but appropriately emphasizes that good data management practices should not end at fielding; even data generated during the Operating and Support (O&S) phase can be used to inform doctrine and future acquisitions.

The TEMP should document requirements for a common T&E database as early as possible, ideally starting at Milestone A, and be updated as needed. As the long-term data needs of the program change, changes to resourcing or other requirements may be necessary in subsequent TEMP updates. The TEMP should specify who is responsible for creating and maintaining the common T&E database, procedures for access, and any special instructions for data storage (e.g., where large or

highly classified data will be stored). Database structure should be documented in the TEMP to the extent that it is known early in a program's life cycle and updated as the implementation matures.

Metadata, Metadata, Metadata

Metadata, or "data about data," is critical to encouraging the reuse of T&E data sets. It enables discovery, aids in understanding and establishing pedigree, helps keep data accessible, and aids in the measurement of overall data quality. Metadata can be broken down into two types: technical metadata and business metadata (Vetterli, Vaduva and Staudt 2000). Technical metadata are perhaps more administrative in nature, and typically include details such as keywords (or "tags"), file names, authors, modification dates, field names, or number of records.

The NCDS and FAIR use the term "discovery metadata." Discovery metadata include technical metadata and describe certain properties of the data set that help narrow a search. These properties may be broad (e.g., creator, contributor, date modified, keywords, or version number) or program-specific (e.g., block number, software configuration, or test event name). Although discovery metadata will vary from program to program, starting with a common standard when creating discovery metadata, such as the Schema.org Data set markup or the World Wide Web Consortium's (W3C) Data Catalog Vocabulary (DCAT), will improve interoperability.

Conversely, so-called "business" metadata are useful to end users and detail context, assumptions, and limitations of the data set. By describing the data collection process, test conditions, and the system under test, business metadata establish pedigree (Snee and Hoerl 2012) and allow end users to determine whether the data set is appropriate for their use. For T&E data, business metadata may include (but are not limited to):

- A thorough description of the system, including versions of software, which subsystems are present, and how those subsystems communicate.
- A detailed test plan that discusses how the system is being tested and why.
- A detailed description of test execution, to include any deviations from the test plan.

Significant business metadata are readily available, since test agencies already generate test plans, test execution reports, and analyses. In some cases, making a data set reusable may be as simple as combining these documents with the raw data used in the analysis (and a data dictionary to interpret the raw data).

There is no "one-size-fits-all" list of required technical and business metadata elements for T&E data sets,

but minimum information checklists can ensure that the most important metadata elements are present when one is adding a data set to a common T&E database. The Minimum Information for Biological and Biomedical Investigations (MIBBI) project and FAIR-sharing are two examples of repositories of minimum information checklists used in academic fields (Taylor 2008, Standards n.d.). Currently, no corollary for these minimum information checklists exists in the T&E community. Programs should create their own minimum standards for metadata generated during tests, as each program's needs will vary.

Store Raw or Minimally Processed Data

Data stored in a common T&E database should remain in raw or minimally processed form to support reproducibility or verification. If an analysis report with summary statistics is stored in the database, the raw data used to generate those statistics should also be included where possible (as a separate record and linked via unique persistent identifier).

Raw data should be formatted in open standards or machine- and human-readable formats with instructions for how to read the data. Data generated in proprietary, platform-specific, or non-standard formats might not be shareable in raw or human-readable format. In such cases, data should be minimally processed and shared along with documentation detailing the processing steps taken. Alternatively, software or instructions for extracting usable data should be included.

Have a Data Life Cycle Plan

The cost of storing data is low and getting lower every day, but data should not be kept forever as T&E data will lose some utility over time. During a long acquisition cycle, a system may mature (or change in other significant ways) and old data may no longer reflect the system's true performance or reliability. Rather than enacting a generic policy of discarding test data after a pre-defined period of two to five years, for example, programs should consider a more nuanced approach to disposal (or archival) that considers the quality of the data in the common T&E database and their expected remaining value.

Continuous maintenance of data repositories to cull data that have lost value is important. Metadata (specifically, "technical" metadata) can be useful in this endeavor, as it can be used to determine levels of different data quality dimension—i.e., accuracy, timeliness, consistency, and completeness.

For example, timeliness can be decomposed into currency, or the length of time since a record's last

update, and volatility, which describes the frequency of updates (Wand and Wang 1996). Both currency and volatility are measurable, and database managers can track them over time, contributing to an increased understanding of whether that data set is still acceptable for use. A data set that has not been accessed, modified, or cited in an analysis for more than five years may no longer be acceptable for use in a particular evaluation, whereas a more current data set might be. The totality of technical metadata in a common T&E database can be used to assess overall data quality at the program level and to understand different “depreciation rates” for different types of data.

Summary

Because the value of T&E data sets is not immediately depleted when we process them for the first time, reusing them multiple times throughout a program’s life cycle is one way to extract as much knowledge as possible. Extensive and formalized reuse of T&E data sets may even shift system evaluations left in a program’s life cycle, reduce or eliminate redundancies associated with collection of data that may exist elsewhere, increase the transparency and credibility of analyses, encourage innovation, and allow us to perform richer analyses. Continued leadership at the Service level to align T&E data management policy with strong Department-level guidance will encourage more consistent reuse of data sets across the T&E community. □

ANDREW FLACK is a Research Associate at the Institute for Defense Analyses in Alexandria, VA. His work at IDA focuses on data management, reproducible research techniques, data visualization, and predictive analytics. He holds a B.S. in Biomechanics from the University of Rochester, and an M.S. in Data Analytics Engineering from George Mason University. Email: afleck@ida.org.

References

- 2004. “Data Sharing in a Net-Centric Department of Defense.” *DoD Directive 8320.02*. December 2.
- n.d. “Dataset.” *Schema.org*. Accessed February 2019. <https://schema.org/Dataset>.
- Defense Acquisition University. 2017. “Defense Acquisition Guidebook.”
- Director, Operational Test and Evaluation (DOT&E). 2010. “Guidance on the Use of Design of Experiments (DOE).”
- DoD Chief Information Officer (CIO). 2003. “Net-Centric Data Strategy.” March 9.
- . 2007. “Net-Centric Services Strategy.” March.
- 1991. “DoD Data Administration.” *DoD Directive 8320.01*. September 26.
- Google. n.d. “Google Developers Reference.” *Dataset*. Accessed February 2019. <https://developers.google.com/search/docs/data-types/dataset>.
- Government Linked Data Working Group. 2014. “W3C Recommendation.” *Data Catalog Vocabulary (DCAT)*. January 16. Accessed February 2019. <https://www.w3.org/TR/vocab-dcat/>.
- 2006. “Guidance for Implementing Net-Centric Data Sharing.” *DoD 8320.02-G*. April 12.
- 2015. “Implementing the Sharing of Data, Information, and Information Technology (IT) Services in the Department of Defense.” *DoD Instruction 8320.07*. August 3.
- Mayer-Schönberger, Viktor, and Kenneth Cukier. 2013. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. New York: Houghton Mifflin Harcourt.
- 2015. “Operation of the Defense Acquisition System.” *DoD Instruction 5000.02*. January 7.
- Pasquetto, Irene V., Bernadette M Randles, and Christine L Borgman. 2017. “On the Reuse of Scientific Data.” *Data Science Journal* 16 (8): 1-9. doi:<https://doi.org/10.5334/dsj2017-008>.
- 2013. “Sharing Data, Information, and Information Technology (IT) Services in the Department of Defense.” *DoD Instruction 8320.02*. August 5.
- Snee, Ronald D., and Roger W. Hoerl. 2012. “Inquiry on Pedigree: Do you know the quality and origin of your data?” *Quality Progress*: 66-68.
- n.d. “Standards.” *FAIRSharing*. Accessed February 2019. <https://www.FAIRsharing.org>.
- Taylor, Chris F., et al. 2008. “Promoting Coherent Minimum Reporting Guidelines for Biological and Bio-medical Investigations: the MIBBI Project.” *Nature Biotechnology* 26 (8): 889-896. doi:[10.1038/nbt.1411](https://doi.org/10.1038/nbt.1411).
- Vetterli, Thomas, Anca Vaduva, and Martin Staudt. 2000. “Metadata Standards for Data Warehousing: Open Information Model vs. Common Warehouse Metadata.” *Special Interest Group on Management of Data (SIGMOD) Record (Association for Computing Machinery)* 29 (3): 68-75. doi:[10.1145/362084.362138](https://doi.org/10.1145/362084.362138).
- Wand, Yair, and Richard Y. Wang. 1996. “Anchoring Data Quality Dimensions in Ontological Foundations.” *Communications of the ACM (Association for Computing Machinery)* 39 (11): 86-95. doi:[10.1145/240455.240479](https://doi.org/10.1145/240455.240479).
- Wilkinson, et al. 2016. “The Fair Guiding Principles For Scientific Data Management and Stewardship.” *Scientific Data*.