



OPERATIONS RESEARCH CENTER
United States Military Academy
West Point, New York 10996

21 May 2021

Technical Report

USMA 2021 Corrosion Data Analysis Study: Assessing Maintenance Data Quality for H-60 and HMMWV Platforms

Prepared By

MAJ Raymond Vetter
Department of Systems Engineering
United States Military Academy
West Point, NY

Prepared For

Office of Corrosion Policy and Oversight
Under-Secretary of Defense for Acquisition and
Sustainment
Alexandria, VA



DISTRIBUTION A. Approved for public release; distribution is unlimited

Report: ORCEN-TR-2102

The views and opinions expressed or implied in this publication are solely those of the author(s) and should not be construed as policy or carrying the official sanction of the United States Military Academy, the Department of Army, the Department of Defense, or other agencies or departments of the US government.



Abstract

Corrosion is estimated to account for over 20 billion dollars' worth of maintenance parts and labor each year in the Department of Defense (DoD). The DoD's Corrosion Policy and Oversight (CPO) directorate is charged with developing and implementing a long-term strategy aimed to improve military readiness and operational efficiencies for DoD equipment and infrastructure through effective corrosion prevention and control technologies. The Maintenance and Availability Data Warehouse (MADW) was developed to improve understanding of the corrosion problem, seeking to inform corrosion policy and funding decisions. However, MADW data relies on raw maintenance data collected by the Services at the organizational and depot levels. For CPO to make data-informed decisions to effectively mitigate and manage corrosion across the Services, it must first have high quality data from users and maintainers. CPO seeks to understand the maintenance data quality for the H-60 Blackhawk, which is an aviation platform common to the Army, Navy, and Air Force. This work focuses on data completeness and consistency, using approximate string matching to assign scores to each maintenance record. The Jaro-Winkler similarity algorithm was selected to assess the closeness of match between string data fields for maintenance entries. Pairwise comparison and assignment of Jaro-Winkler similarity scores highlight whether maintenance records are consistent between the coded data fields and the free text narratives where higher scores indicate greater consistency across records. The use of the Jaro-Winkler similarity algorithm is also used to assess whether a maintenance record is corrosion-related. This comparison utilizes discrepancy codes, free text fields, and defined corrosion key terms, as defined by each Service. A similar assessment is performed on the High Mobility Multipurpose Wheeled Vehicle (HMMWV) to assess if this approximate string matching technique is extendable to ground platforms. This assessment highlights the quality of the free text fields and indicates if corrosion-coded maintenance records may be mislabeled. The analysis then informs current best practices for maintenance data across the Services and recommends potential improvements for data collection methods.

Key Words: Approximate String Matching; Corrosion; Data Analysis; Data Quality; Expert Systems



CUI

2021 ORCEN CPO Technical Report
Vetter

This Page Intentionally Left Blank

CUI



Executive Summary

The DoD's Corrosion Policy and Oversight (CPO) directorate's mission is to develop and implement a long-term strategy aimed to improve military readiness and operational efficiencies for DoD equipment and infrastructure through effective corrosion prevention and control technologies. For CPO to make data-informed decisions to effectively mitigate and manage corrosion across the Services, it must first have high quality data from maintainers. CPO seeks to understand the maintenance data quality for the H-60 Blackhawk and the High Mobility Multipurpose Wheeled Vehicle (HMMWV). This study seeks to help CPO better answer two questions relating to H-60 and HMMWV maintenance data: (1) "How accurate are the corrosion estimates?"; (2) "How much can I rely on this data?". This study provides an assessment of maintenance data quality and recommended best practices for consideration. It also proposes additional systems for further analysis.

Research Objectives

To address the problems stated above, this study examines the H-60 Blackhawk (Navy and Air Force) and HMMWV (Army and Marine Corps). The methodology is flexible and can be extended to additional platforms. The study began with the following research objectives to address the problems stated above.

- *Objective 1:* Assess data quality for Navy and Air Force H-60 rotary wing aircraft maintenance data.
- *Objective 2:* Identify best practices and provide recommendations from the Services regarding corrosion-related maintenance data collection
- *Objective 3:* Adapt data quality analysis to the HMMWV to determine scalability of the methodology to a cross-Service ground system.
- *Objective 4:* Recommend additional ground systems for future analysis.

Methodology

To address these objectives, the analyst developed an approach that seeks to assess data consistency and data completeness for the maintenance datasets. Approximate string matching compared the "coded" data fields to the corresponding "free text" data fields in order to determine if the records were consistent. The MADW Corrosion Dictionary was incorporated to assess whether free text content was corrosion-related or not. Pairing this evaluation of the free text content with the coded fields allowed for evaluation as to whether coded and free text fields were correctly classified as corrosion-related or not.

Results

When assessing H-60 maintenance records, no single record type displayed dominance in terms of having the highest quality data records. For the object receiving maintenance, the Air Force H-60 depot-level maintenance records scored the highest, with a 96.19% object match consistency score. For maintenance action taken, Navy H-60 organizational maintenance records scored the highest, with an action taken match consistency score of 84.05%. For maintenance discrepancy, Navy H-60 organizational maintenance records again scored the highest, with a malfunction match consistency score of 78.52%. Navy and Air Force organizational and depot-level maintenance records exhibit high corrosion match accuracy levels (all exceed 83%). Object data fields had the lowest record completeness for both Services. This is likely because part codes are manually entered and are five to eight characters in length and there are several thousand available codes, which introduces more opportunity for input errors from the users.



Record Type	Corrosion Match Accuracy	Completeness
Navy Organizational H-60 Maintenance Data	83.60%	88.05% WUCs complete
Navy Depot H-60 Maintenance Data	92.66%*	No AT Codes available in records
Air Force Organizational H-60 Maintenance Data	96.57%	98.31% WUCs complete
Air Force Depot H-60 Maintenance Data	88.11%	98.66% MALCODES complete

When assessing the Army and Marine Corps HMMWV organizational maintenance records, the corrosion match accuracy scores were slightly lower than the H-60 maintenance records. However, when further evaluating the HMMWV records, recall scores for Army and Marine Corps were very low (1.45% and 25.75% respectively). This indicates that HMMWV records exhibit a high percentage of records have 'corrosion-positive' text in the narrative description but are assigned a 'corrosion-negative' Failure/Defect Code. The analysis also determined that Army records are overly reliant on generic malfunction codes (70.12% of codes are '099' – OTHER/UNKNOWN/BROKEN).

Record Type	Corrosion Match Accuracy	Completeness
Army Organizational HMMWV Maintenance Data	74.83%	99.06% malfunctions complete
Marine Corps Organizational HMMWV Maintenance Data	82.03%	No Part Codes or AT Codes available in records

Findings

Best Practices:

1. Feedback loops improve maintenance actions (Navy ADDR Program) and maintenance data quality (Air Force DIT).
2. Specific corrosion-related malfunction codes increase granularity of a specific discrepancy.
3. Engaged unit-level leadership in maintenance organizations increases maintainer proficiency.

Recommendations:

1. Require mandatory data fields to improve data completeness.
2. Incorporate Data Integrity Teams (or similar) to provide data feedback at organizational level.
3. Standardize data collection and processes to decrease variability in maintenance narratives.
4. Standardize coded data fields across Services and platforms to allow tools to be transferable.
5. Improve data collection at point of entry with Natural Language Processing tools.

Systems for Further Analysis:

1. Joint Light Tactical Vehicle (JLTV)
2. High Mobility Artillery Rocket System (HIMARS)
3. Multiband Multi Mission Radio (MBMMR)
4. M4 Carbine
5. M2 .50 Caliber Machine Gun

This study highlights that Navy and Air Force H-60 maintenance records are of high quality. The best practices and recommendations from the H-60 maintenance data analysis highlights the need for feedback loops, engaged small unit leaders, mandatory fields to reduce missing information, additional training and standardization for maintainers, and AI tools at the point of entry. Based on the Army and Marine Corps HMMWV maintenance records analysis, HMMWVs have lower maintenance data quality than aviation systems. While the overall corrosion accuracy of HMMWV maintenance data appears almost comparable to the H-60 accuracy scores, a significant percentage of records with corrosion-related words are not tagged with a corrosion-related malfunction code. This indicates a need for further standardization and training across the maintenance workforce.



CUI

This Page Intentionally Left Blank

CUI



About Us

Established in 1988 inside the Department of Systems Engineering (DSE), the Operations Research Center (ORCEN) provides USMA with a dedicated analytical capability that routinely tackles problems of national significance for the Army and the Department of Defense (DoD). Leveraging the systems engineering and mathematical modeling skills of a small cadre of full-time analysts, the ORCEN executes an annual research portfolio of six to eight reimbursable projects across varied domains, including (but not limited to) operations assessment, human resource management, geospatial modeling, maintenance process improvement, and force structure decisions.

Immaterial of the domain, when the ORCEN analyst's abilities and the project sponsor's needs properly align, history has shown that the results can be quite good. Specifically, since 2005 the ORCEN has won the Dr. Wilbur B. Payne Memorial Award for Excellence in Analysis five times – an annual award, given under the signature of the Secretary of the Army, which recognizes the best operations research work that makes a significant contribution to the force. In addition to providing invaluable insights to their sponsors, ORCEN projects enrich cadet education, professionally develop Operations Research / Systems Analysis (ORSA) faculty, integrate emerging technologies into the Academic Program, and sustain ties between the Academy and the Army.

Beyond executing year-long, faculty-led research projects, the ORCEN research often extends into the classroom through ORCEN aligned cadet capstone experiences or indirectly as junior faculty return to the classroom with their research experience. Often ORCEN projects are accompanied with cadet capstone research or summer internships through the Academic Individual Advanced Development (AIAD) program. Cadets and analysts will spend three to four weeks between May and August embedded with an organization and upon return to USMA they will continue their research as part of the Department's capstone program. As junior faculty depart the ORCEN and return to the classroom, they will bring portions of their research projects with them as in-class examples to demonstrate to the cadets how the material they learn in DSE's courses directly applies to real-world problems they may encounter.

For more information regarding the ORCEN's reimbursable research or cadet capstone support please contact:



Operations Research Center
Attn: LTC(P) James R. Enos, PhD, Director
Department of Systems Engineering
Mahan Hall, Bldg. 752
West Point, NY 10996
james.enos@westpoint.edu
845-938-3688



CUI

This Page Intentionally Left Blank

CUI



Table of Contents

Section 1 Introduction and Organization	1
1.1. Introduction	1
1.2. Project Overview	1
1.3. Organization	1
Section 2 Background	2
2.1. Introduction	2
2.2. MADW Overview	3
2.3. Army Maintenance Data Collection Process	4
2.4. Navy Maintenance Data Collection Process	5
2.5. Air Force Maintenance Data Collection Process	6
2.6. Problem Statement	7
2.7. Literature Review	7
2.7.1. Data Quality	7
2.7.2. Record Linkage & Approximate String Matching	8
2.8. Hypotheses	11
2.8.1. Hypothesis 1 – Navy H-60	11
2.8.2. Hypothesis 2 – Air Force H-60	11
2.8.3. Hypothesis 3 – Army HMMWV	12
2.8.4. Hypothesis 4 – Marine Corps HMMWV	12
2.9. Study Limitations	12
Section 3 Approach and Methodology	14
3.1. General Approach	14
3.2. Objective 1: Maintenance Data Quality for Navy & Air Force H-60	15
3.2.1. Data Sample	15
3.2.2. Research Procedures	19
3.3. Objective 3: Maintenance Data Quality for HMMWV	24
3.3.1. Data Sample	24
3.3.2. Research Procedures	26
3.4. Assumptions	27
Section 4 Findings	28
4.1. Findings: Navy and Air Force H-60 Maintenance Data Quality	28



4.2. Findings: Best Practices and Recommendations for Corrosion Maintenance	38
4.3. Findings: Army and Marine Corps HMMWV Maintenance Data Quality	41
4.4. Findings: Recommended Ground Systems for Future Study	49
Section 5 Conclusions and Implications	51
5.1. Summary of Findings	51
5.2. Impact of the Study	54
5.3. Recommended Future Work	55
Appendix A References	56
Appendix B Code Used to Assess H-60 and HMMWV Maintenance Data	59
Appendix C MADW-provided Corrosion Search Terms	60
Appendix D Navy Corrosion-Related MALCODES	63
Appendix E Air Force Corrosion-Related MALCODES	65
Appendix F H-60 Coded Field to Free Text Comparison	66
Appendix G ADDR Process Flow Chart	82
Appendix H HMMWV Coded Field to Free Text Comparison	83



List of Figures

Figure 1: MADW Data Sources [10, p. 3].....	3
Figure 2: Data Quality Framework [16, p. 304].....	8
Figure 3: General Linkage Process [20, p. 948]	9
Figure 4: Study Methodology	15
Figure 5: Record Linkage Procedure for H-60 Maintenance Data	20
Figure 6: Record Linkage for Corrosion-Related Maintenance Records	22
Figure 7: Error Matrix for Corrosion-Related Maintenance Records	23
Figure 8: Record Linkage Procedure for HMMWV Maintenance Data	26
Figure 9: Navy H-60 Organizational Maintenance Data Part Code to Free Text Comparison	28
Figure 10: Navy H-60 Organizational Maintenance Data Action Taken Code to Free Text Comparison ..	29
Figure 11: Navy H-60 Organizational Maintenance Data Malfunction Code to Free Text Comparison	30
Figure 12: Navy H-60 Organizational Corrosion-Positive Malfunction Code	31
Figure 13: Navy H-60 Organizational Corrosion-Negative Malfunction Code to Free Text Comparison ...	32
Figure 14: Navy H-60 Organizational Corrosion-Consistency Error Matrix and Performance Metrics.....	33
Figure 15: Army HMMWV Organizational Maintenance Data Part Code to Free Text Comparison	42
Figure 16: Army HMMWV Organizational Maintenance Data Action Taken Code.....	43
Figure 17: Army HMMWV Organizational Maintenance Data Malfunction Code	44
Figure 18: Army HMMWV Organizational Corrosion-Positive Malfunction Code	45
Figure 19: Army HMMWV Organizational Corrosion-Negative Malfunction Code	46
Figure 20: Army HMMWV Organizational Corrosion-Consistency Error Matrix	47
Figure 21: Navy H-60 Depot Maintenance Data Part Code to Free Text Comparison.....	66
Figure 22: Navy H-60 Depot Maintenance Data Malfunction Code to Free Text Comparison.....	67
Figure 23: Air Force H-60 Organizational Maintenance Data Part Code to Free Text Comparison	68
Figure 24: Air Force H-60 Organizational Maintenance Data Action Code to Free Text Comparison	69
Figure 25: Air Force H-60 Organizational Maintenance Data Malfunction Code to.....	70
Figure 26: Air Force H-60 Depot Maintenance Data Part Code to Free Text Comparison	71
Figure 27: Air Force H-60 Depot Maintenance Data Action Code to Free Text Comparison.....	72
Figure 28: Air Force H-60 Depot Maintenance Data Malfunction Code to Free Text Comparison	73
Figure 29: Navy H-60 Depot Corrosion-Positive Malfunction Code to Free Text Comparison	74
Figure 30: Navy H-60 Depot Corrosion-Consistency Error Matrix.....	75
Figure 31: Air Force H-60 Organizational Corrosion-Positive Malfunction Code to	76



Figure 32: Air Force H-60 Organizational Corrosion-Negative Malfunction Code to.....	77
Figure 33: Air Force H-60 Organizational Corrosion-Consistency Error Matrix.....	78
Figure 34: Air Force H-60 Depot Corrosion-Positive Malfunction Code to Free Text Comparison	79
Figure 35: Air Force H-60 Depot Corrosion-Negative Malfunction Code to Free Text Comparison	80
Figure 36: Air Force H-60 Depot Corrosion-Consistency Error Matrix	81
Figure 37: ADDR Process Flow Chart	82
Figure 38: Marine Corps HMMWV Organizational Maintenance Data Malfunction Code	83
Figure 39: Marine Corps Organizational Corrosion-Positive Malfunction Code	84
Figure 40: Marine Corps Organizational Corrosion-Negative Malfunction Code.....	85
Figure 41: Marine Corps HMMWV Organizational Corrosion-Consistency Error Matrix	86



List of Tables

Table 1: Navy H-60 Organizational Level Maintenance Data Sample	16
Table 2: Navy H-60 Depot Level Maintenance Data Sample	16
Table 3: MADW Navy/Marine Corps Aviation Corrosion Dictionary Sample	17
Table 4: Air Force H-60 Organizational Level Maintenance Data Sample	17
Table 5: Air Force H-60 Depot Level Maintenance Data Sample	18
Table 6: MADW Air Force Corrosion Dictionary Sample	18
Table 7: Classification Categories for J-W Similarity Scores	21
Table 8: MADW Corrosion Poison/Exclusionary Words Sample	23
Table 9: Army Organizational Level Maintenance Data Sample	25
Table 10: Marine Corps Organizational Level Maintenance Data Sample	25
Table 11: MADW Marine Corps & Army Ground Systems Corrosion Dictionary Sample	26
Table 12: Results for H-60 Maintenance Records	38
Table 13: Results for HMMWV Maintenance Records	49
Table 14: Navy and Air Force H-60 Corrosion Consistency and Completeness Summary	52
Table 15: Army and Marine Corps HMMWV Corrosion Consistency and Completeness Summary	54
Table 16: MADW-provided Sample of Navy/Marine Corps Aviation Corrosion Search Terms	60
Table 17: MADW-provided Sample of Air Force Corrosion Search Terms	60
Table 18: MADW-provided Sample of Army Aviation Corrosion Search Terms	61
Table 19: MADW-provided Sample of Army/Marine Corps Ground Systems Corrosion Search Terms	61
Table 20: MADW-provided Exclusionary Terms	62



Acronyms

ADCS	Automated Data Capture System
ADDR	Aircraft Delivery Deficiency Report
AFCPCO	Air Force Corrosion Prevention and Control Office
AFLCMC	Air Force Life Cycle Management Center
AFMC	Air Force Materiel Command
AFTO	Air Force Technical Order
AI	Artificial Intelligence
Ao	Operational Availability
ASB	Aviation Support Battalion
AT	Action Taken
BCT	Brigade Combat Team
C/DA	Cost per Day of Availability
CCDC	Combat Capabilities Development Command
CCPE	Corrosion Control Prevention Executive
CNN	Convolutional Neural Network
CPC	Corrosion Prevention and Control
CPCIPT	Corrosion Prevention and Control Integrated Product Team
CPO	Corrosion Policy and Oversight
DA	Department of the Army
DAC	Data & Analysis Center
DAU	Defense Acquisition University
DECKPLATE	Decision Knowledge Programming for Logistics Analysis and Technical Evaluation
DLM	Depot Level Maintenance
DIT	Data Integrity Teams
DDSN	Document Date and Serial Number
DoD	Department of Defense
E&E	Examiner and Evaluator
FEDLOG	Federal Logistics database
FY	Fiscal Year
GCSS	Global Combat Support System
GCSS-A	Global Combat Support System-Army
GCSS-MC	Global Combat Support System-Marine Corps
HIMARS	High Mobility Artillery Rocket System
HMMWV	High Mobility Multipurpose Wheeled Vehicle
IMDS	Integrated Maintenance Data System
ISO	International Organization for Standardization
IPT	Integrated Product Team
JCN	Job Control Number
JLTV	Joint Light Tactical Vehicle
LMI	Logistics Management Institute
LRC	Logistics and Readiness Center
LSTM	Long Short-Term Memory (unit)
MADW	Maintenance and Availability Data Warehouse
MAF	Maintenance Action Form
MAL	Malfunction
MALCODES	Malfunction Codes



MBMMR	Multiband Multi Mission Radio
MC	Mission Capability
MERC	Mercer Engineering Research Center
ML	Machine Learning
MLRS	Multiple Launch Rocket System
MRR	Maintenance Request Register
NAH	Non-Armored HMMWVs
NAMP	Naval Aviation Maintenance Program
NAVAIR	Naval Air Systems Command
NBNC	Noted But Not Corrected
NLP	Natural Language Processing
OCONUS	Outside the Continental United States
OEM	Original Equipment Manufacturers
OOMA	Optimized Organizational Maintenance Activity
ORCEN	Operations Research Center
OSD	Office of the Secretary of Defense
PMA	Program Management Activity
PMCS	Preventative Maintenance Checks and Services
REMIS	Reliability and Maintainability Information System
RNN	Recurrent Neural Network
SDC	Sample Data Collection
SME	Subject Matter Expert
TAT	Turnaround Time
TO	Technical Order
UAH	Up-Armored HMMWVs
USD	Under-Secretary of Defense
(USD (A&S))	Under-Secretary of Defense (Acquisition and Sustainment)
USMA	United States Military Academy
WO	Work Order
WUC	Work Unit Code



Section 1

Introduction and Organization

1.1. Introduction

This report summarizes the work done by MAJ Raymond Vetter in support of the Corrosion Policy and Oversight (CPO) directorate, Office of the Under-Secretary of Defense (Acquisition and Sustainment) from July 2020 to May 2021. The analysis in this technical report was enabled by the cooperation and assistance of the Logistics Management Institute (LMI), CPO's primary consulting firm. This research and its findings would not have been possible without the assistance of LMI.

1.2. Project Overview

This study supports the CPO's efforts to understand the impacts of corrosion and how maintenance data across services and platforms can be used to make informed decisions at the senior levels of the Department of Defense (DoD). This study has four research objectives. The first objective was to assess the data quality for U.S. Navy and U.S. Air Force H-60 rotary wing aircraft maintenance data. Specifically, the analyst considered the consistency and completeness of data for the services regarding the object receiving maintenance, the maintenance action taken, the maintenance malfunction, and whether a maintenance malfunction was corrosion related. The second objective is to identify best practices from the Services regarding corrosion-related maintenance data collection.

The third objective of the study is to assess data quality for the High Mobility Multipurpose Wheeled Vehicle (HMMWV) to determine scalability of the methodology to a cross-Service ground system. This objective utilized a similar assessment methodology from the first objective. This demonstrated the flexibility of the approach. After assessing the HMMWV maintenance data quality for Army and Marine Corps records, the final objective was to recommend additional ground systems for future analysis.

1.3. Organization

The remainder of this report is organized as follows:

- Section 2 provides background on the problem, current data collection methods, and relevant definitions and established methods to perform data quality analysis.
- Section 3 explains the approach and methodology used to assess data quality for the U.S. Navy and U.S. Air Force H-60 maintenance records as well as for the cross-Service HMMWV records.
- Section 4 provides the analysis findings and results. It also highlights best practices and recommendations for maintenance data collection as well as recommended ground systems for future analysis.
- Section 5 provides the impact and conclusions resulting from the study and possible areas of future work.
- Appendix A contains the references used in the conduct of this research.
- Appendix B contains the R code developed to conduct the analysis.
- Appendix C contains the Maintenance and Availability Data Warehouse (MADW)-provided corrosion search terms and exclusionary terms used to assess corrosion maintenance records.
- Appendix D & E contain corrosion-specific malfunction codes used for the analysis.
- Appendix F contains results and statistics supporting H-60 data quality analysis.
- Appendix G contains the Navy's ADDR process flow chart.
- Appendix H contains results and statistics supporting HMMWV data quality analysis.



Section 2 Background

2.1. Introduction

Corrosion represents a significant challenge for equipment and infrastructure across the Department of Defense (DoD). There are five main areas where corrosion has negative impacts: safety to Servicemembers, military readiness and operational availability, system performance, maintenance costs, and environmental damage [1, p. 1]. These negative effects of corrosion rarely occur in isolation. Rather, corrosion-related issues may contribute to multiple areas. Structural cracking on an aircraft may degrade system performance while simultaneously causing a safety hazard and eventually increasing equipment maintenance costs. Unfortunately, many examples exist where corrosion-related safety hazards have directly led to the loss of Servicemember lives [2, p. 1].

In December 2003, a DoD report to Congress defines corrosion as “the unintended destruction or deterioration of a material due to interaction with the environment. It includes such varied forms as rusting; pitting; galvanic reaction; calcium or other mineral build up; degradation due to ultraviolet light exposure; and mold, mildew, or other decay” [2, pp. 4-5]. Several DoD-sponsored studies have demonstrated the significant cost of corrosion. Direct costs have been estimated to be between \$10 and \$20 billion. Indirect costs, such as the loss of equipment usage and the loss of training time, are not quantified but would increase the total cost of corrosion [2, p. 7].

The Defense Acquisition University (DAU) and Office of the Secretary of Defense (OSD) have also attempted to quantify the impact of corrosion on maintenance costs and readiness. DAU has even higher estimates for the cost of corrosion. DAU estimates that corrosion costs the DoD more than \$20 billion annually and that “approximately 25% of all systems/equipment and facility maintenance is attributable to corrosion” [3]. OSD estimates that “approximately 30% of current DoD corrosion costs could be avoided through investment in sustainment, design, and manufacture and other preventative measures such as paint and avoidance of dissimilar metals” [4].

To address corrosion costs, Congress enacted Public Law 107-134 Section 1067 (codified in 10 U.S.C. 2228) titled “Prevention and mitigation of corrosion of military infrastructure and equipment” which implemented annual reporting by Services and DoD [3]. This legislation led to the creation of the Office of Corrosion Policy and Oversight (CPO) [5, pp. 1-2]. CPO is responsible for developing a long-term DoD strategic plan to implement standards and improve strategies to prevent, detect, and treat corrosion [5, p. 6]. Prior to August 2018, CPO reported directly to the Under Secretary of Defense for Acquisition, Technology, and Logistics. DoD restructuring led to the creation of the Under Secretary of Defense for Acquisition and Sustainment (USD (A&S)). As of August 1, 2018, DoD relocated CPO to the Office of the USD (A&S). More specifically, the Deputy Assistant Secretary of Defense for Materiel Readiness oversees CPO [1, pp. 4-5]. General responsibility for DoD corrosion prevention and control (CPC) rests with the Under Secretary of Defense for Acquisition & Sustainment, (USD (A&S)), which establishes policy, administers research and funding, and operates the CPO [3].

A few key actors support CPO in its effort to develop and implement a long-term strategy aimed to improve military readiness and operational efficiencies for DoD equipment and infrastructure through effective corrosion prevention and control technologies. Each Service has a Corrosion Control Prevention Executive (CCPE). The CCPE is responsible for coordinating corrosion prevention and mitigation activities within its respective military department within OSD [5, p. 2]. Logistics Management Institute (LMI) has been CPO’s primary consulting firm. They provide both scientific expertise and data management capabilities to CPO to help solve corrosion-related problems. For example, LMI helped develop the DoD’s methodology to

quantify corrosion cost and impact, which was ultimately approved by the corrosion prevention and control integrated product team (CPCIPT).

2.2. MADW Overview

The Maintenance and Availability Data Warehouse (MADW) is a data repository and management system designed and managed by LMI. MADW contains availability, cost, inventory, and transactional data on nearly every weapons system and readiness reportable piece of equipment in the DoD [6]. MADW ingests 46 authoritative data systems monthly and contains over 1.1 billion labor transactions and 500 million supply transactions dating back to 2005. It was created to help understand the impact of corrosion of DoD weapon systems, infrastructure, and facilities [7]. MADW's overall goal for sustainment is to ensure every weapon system is at or above target availability at the lowest cost [8]. To measure sustainment success, three metrics capture weapon system effectiveness and efficiency:

- Operational Availability (Ao) – effectiveness metric calculated by dividing the total asset inventory by the available inventory. Ao is measured as a percentage of inventory available.
- Mission Capability (MC) – quantifies the inventory assigned to operational units. To calculate MC primary mission asset inventory is divided by the available inventory. MC is measured as a percentage of inventory available for operational units.
- Cost per Day of Availability (C/DA) – effectiveness metric that divides the total achieved available days by the total maintenance costs. C/DA is measured in terms of cost due to corrosion per day of equipment availability [9].

MADW ingests, cleans, and processes data from multiple data sources (see Figure 1). Top-down sources provide cost and availability information. Top-down sources include depot and original equipment manufacturer (OEM) information. Bottom-up records includes material and labor transactional data from the organizational/unit/field level. Machine Learning (ML) algorithms process records to solve for maintenance actions and objects at the task level [6].



Figure 1: MADW Data Sources [10, p. 3]



MADW is invaluable to CPO and the CCPEs to make informed decisions about how best to manage corrosion efforts; however, MADW still relies on data from the Services. The organizational and depot data quality directly impacts the cleaned and refined dashboards and analytics that MADW provides. Each Service collects maintenance data differently. Summarizing the bottom-up data collection process for each Service establishes a shared understanding and underscores differences across the Services.

2.3. Army Maintenance Data Collection Process

Within the Army, the maintenance data collection process differs significantly between ground and aviation platforms. Both aviation and ground platforms have field level maintenance and depot level maintenance. Field level maintenance is performed internal to a unit (battalion or brigade), which depot maintenance is performed at a consolidated location separate from any specific Army unit. Depot maintenance is more intrusive and time-intensive than field maintenance, but periodic depot maintenance extends the lifecycle of platforms. It is instructive to show how the maintenance data is collected from Army bottom-up resources. In Army ground systems units, this is the maintenance process:

1. Soldiers, the equipment operators, conduct preventative maintenance checks and services (PMCS) on the motorpool line in their units and record faults on DA5988's, the Equipment Maintenance and Inspection Worksheet.
2. The operators then get a mechanic, also usually a Soldier, to verify or correct faults from the DA5988. If the maintenance work requires parts to be ordered or significant labor, this is recorded on a DA2405, the Maintenance Request Register (MRR). The DA2405 (or 5989E in its electronic version) is used at both unit and support levels used to record all work requests received and handled by maintenance activities. It is supposed to capture cost of parts, hours of labor, descriptions of the work completed, and the objects that received the maintenance. The maintenance actions taken are summarized by the mechanic directly on the DA2405 for later digitizing of the document or are dictated to a clerk for input into the electronic version of the MRR [10] [11].
3. Parts are ordered and when on-hand, the work is completed and logged on the DA2405, closing out the maintenance action.
4. The digitized data from the DA2405 is uploaded through the Global Combat Support System – Army (GCSS-A), which is the data repository for maintenance and supply chain actions in the Army.
5. The MADW then imports the data from GCSS-A [10, p. 4].

The maintenance data collection forms (DA2405 or DA5988E) are filled out by hand or manually in a computer by Soldiers and maintainers, taking free text input data. Free text data is unstructured and highly variable, but often includes the most important information recorded on the maintenance record.

Additionally, in the name of efficiency at the unit level, it is typically a clerk that inputs maintenance data as recorded on the maintenance forms or as dictated by maintainers, meaning that the person who performed the maintenance or observed the work is not necessarily inputting the descriptions into the data system. As a result, many details can be lost in the data collection process, as well as introduce shorthand, acronyms, and slang terms into the data system [10, p. 4].

Alternatively, Army aviation units have an improved data collection process over Army ground units. In general, aviation units have more experienced and trained maintainers that specialize in aircraft maintenance compared to their general-purpose ground maintainer counterparts. Most significantly, the aviation industry is highly regulated, with increased logging, maintenance, and safety requirements, and as a result the data automation efforts have been a focal point for much longer than in Army ground units. In



addition to the Army ground maintenance procedures, aviation units record all airframe usage on the DA2408-12, which is the Aviator's Flight Record logbook. Any time a part is removed, repaired, installed, or otherwise gained or lost from an aircraft, the action is recorded on the DA2410, the Component Removal/Repair/Install/Gain/Loss Record. Aviation units also use the DA2408-13-1 instead of the DA2405, but it serves the same purpose. The MADW primarily imports maintenance data from DA2408-13-1 records, but also has access to DA2408-12 and DA2410 records. Therefore, the bulk of MADW Army aviation data still comes from the free text entry portions of the DA2408-13-1 form [10, p. 5].

2.4. Navy Maintenance Data Collection Process

Like the Army, the Navy has different processes and regulations that guide the collection of aviation data and ground vehicle data. The Marine Corps is the primary user of ground vehicles within the Navy, and Marine Corps Tactical Publication (MCTP) 3-40E – Maintenance Operations, governs maintenance processes and standards in the Marine Corps for ground equipment. The collection of Naval Air Systems Command (NAVAIR) maintenance data is governed by COMNAVAIRFORINST 4790.2C – The Naval Aviation Maintenance Program (NAMP).

The Navy has three levels of maintenance: organizational level (O-level), intermediate level (I-level), and depot level (D-level). O-level maintenance focuses on inspecting, servicing, lubricating, adjusting, and replacing parts, minor assemblies, and subassemblies of a system. O-level maintenance is typically performed by the unit that owns the system. The Optimized Organizational Maintenance Activity (OOMA) collects O-level maintenance actions via work orders. I-level maintenance organizations are collocated with the owning unit of a weapon system but may also support other units. I-level maintenance typically consists of testing and repair of components and support equipment, calibration, technical assistance to supported units, and among other specified I-level maintenance tasks. D-level maintenance occurs at specific, consolidated locations and consists of major maintenance activities. D-level maintenance activities are collected in the Automated Data Capture System (ADCS). Typical D-level activities include major overhaul or rebuilding of engines, parts, assemblies, subassemblies, and end items, calibration by Navy laboratories, modification of engines and support equipment, and technical and engineering assistance [12, pp. 3-1-2]. The Marine Corps maintenance levels for ground vehicles are similar but O-level and I-level maintenance are grouped within the "field level maintenance" category [13, p. iii].

For Marine Corps and Navy O-level maintenance, upon completion of a mission, the operator initiates a maintenance action form (MAF) or work order (WO) (Form SWP 4826) for each discrepancy. For discrepancies discovered by other than user or crew, the MAF or WO will be initiated by the person who discovered the discrepancy. In the case of When Discovered Code 'O' (Administrative), Maintenance Control will initiate the MAF or WO. Corrosion Prevention MAF or WOs may be initiated by any operator, crew, or maintenance personnel. OOMA prompts the user for required data fields during MAF or WO initiation. A job control number (JCN) is automatically assigned when the MAF or WO is approved. The Type MAF or WO Code, received date, received time, and several administrative fields related to the system are pre-filled. The received date and time can be changed. Work center, discrepancy, initiator, and up or down status field shall be filled in prior to saving to the database.

Upon reviewing MAF or WOs, Maintenance Control has the option to modify all fields of the MAF or WO. Upon MAF or WO approval, the MAF or WO is ready to be printed. Maintenance Control prints a two-part MAF or WO. Once the MAF or WO is printed the original copy is placed on the right side of the system logbook and shall remain as long as the discrepancy remains outstanding. A copy is routed to the



appropriate work center. Work centers shall retain the copy until it appears on the next Work Center Work Load Report.

When the maintainer completes the corrective action, Maintenance Control reviews, approves, or rejects MAF or WOs. Upon approval of MAF or WO completion, Maintenance Control prints a two-part MAF or WO. The original completed copy is then placed on the left side of the ADB where it shall remain for 10 subsequent flights following the completion date. The outstanding copy is removed from the right side of the ADB and discarded. The completed copy is retained for historical files.

When a corrective action requires parts or materials, the Maintenance Control Supervisor will enter the appropriate project code and priority designator on the MAF or WO, using the project and priority assignment online process. The MAF or WO is electronically forwarded to Material Control's document date and serial number (DDSN) assignment online process [12, pp. 15-10-11].

2.5. Air Force Maintenance Data Collection Process

The Air Force maintenance program is governed primarily by Air Force Instruction 21-101: Aircraft and Equipment Maintenance Management. The Air Force data collection process is codified in Technical Order (TO) 00-20-2: Maintenance Data Documentation.

The Air Force has three levels of maintenance: organizational level (O-level), intermediate level (I-level), and depot level (D-level). O-level maintenance focuses on those on-equipment tasks normally performed using the resources of an operating command at an operating location. I-level maintenance consists of those off-equipment tasks normally performed using the resources of the operating command at an operating location or at a centralized intermediate repair facility. D-level maintenance provides the capability to maintain materiel coded for organizational, intermediate and depot levels of maintenance. As the highest level of maintenance, D-level maintenance tasks include system overhaul, upgrading, or rebuilding of parts, assemblies, or subassemblies, and the testing and reclamation of equipment as necessary according to Air Force regulations [14, pp. 323-326].

Air Force data collection at the unit level commences when an operator or maintainer completes an Air Force Technical Order (AFTO) Form 349 – Maintenance Data Documentation Record. This form captures essential information about the aircraft, the discrepancy, the action required, and when the issue was discovered. AFTO Form 349 also allows the operator/maintainer to write a free text narrative for the discrepancy and the corrective action. The maintainer (or another clerk within the unit) then enters the maintenance data from the AFTO Form 349 into the Integrated Maintenance Data System (IMDS). IMDS is the management information system for mobility aircraft, and G081 is the maintenance information system for heavy logistic aircraft (C-40, C-5, C-130, KC-10, KC-135, and C-17). Work center supervisors are responsible for validating the data submitted into IMDS/Web G081. Maintenance data is then transferred to the Reliability and Maintainability Information System (REMIS), which is the Air Force's authoritative enterprise system for all O-level through D-level maintenance information [14] [15]. Air Force maintenance units all have established Data Integrity Teams (DITs) that are responsible for ensuring the unit has complete and accurate data. These DITs identify and quantify problems within their unit and attempt to identify the root cause of the poor data integrity. DITs send error reports to the respective work centers to help educate maintainers and operators that have repeated data quality issues. Units may also implement training programs to correct data integrity trends [14].



2.6. Problem Statement

CPO relies on the MADW to consolidate, process, and assess maintenance and financial data from across the Services. MADW's many algorithms are essential to providing insights about the current state of corrosion and maintenance. When decision makers use maintenance data, their frequently asked questions include:

- "How accurate are the corrosion estimates?"
- "How much can I rely on this data?"

While the MADW serves as a consolidated data source across the Services, the data it ingests comes from the individual Services. As such, the MADW data quality is only as good as the maintenance and financial data provided by each Service. The ORCEN FY20 study for CPO, conducted by MAJ John Case, assessed the quality of Army H-60 maintenance data. The first objective of this study is to assess the quality of maintenance data for the Navy H-60 platform and Air Force H-60 platform. By assessing a common platform during the same timeframe as the Army H-60 (calendar year 2017), conclusions about maintenance data quality will emerge. After assessing maintenance data quality, the second objective is to identify best practices for data collection across the Services. This will rely on the data quality assessment and understanding the data collection processes for the different Services.

After assessing maintenance data quality for the H-60 platform, CPO is interested in whether this methodology is scalable to ground vehicles across the Services. The HMMWV is a reportable vehicle for the Army and the Marine Corps, meaning that the vehicle impact unit readiness statistics for those Services. The study's third objective is to assess the maintenance data quality for the Army and Marine Corps using the same methodology as for the H-60. After conducting this evaluation, the final study objective is to recommend future ground systems to analyze, and the refined research questions follow:

Given that CPO needs to understand to what degree it can rely on MADW data, (Objective 1) is the incoming maintenance data from the Services of high quality for the H-60 Blackhawk? (Objective 2) Having assessed maintenance data for the H-60, what are best practices that emerge? (Objective 3) How does this data quality assessment methodology transfer to ground systems, specifically the HMMWV? (Objective 4) What other ground systems are worthy of further analysis?

2.7. Literature Review

To address the study's questions, two data analysis fields are especially pertinent to the defined objectives:

1. Objective 1: Data Quality
2. Objective 2: Record Linkage and Approximate String Matching

2.7.1. Data Quality

While many definitions for quality data exist, it is important for this study to establish a standard definition of data quality that can apply to maintenance data from different Services and different systems.

A comprehensive survey of data quality dimensions was conducted in 2012 that identifies the components of data quality and assesses which are most widely accepted across the decision making and data science communities. High quality data is critical to improving operations and informing decisions in all high-performing organizations. According to quality management experts, "data quality is appropriate for use...

and meet[s] user needs” [16, p. 301]. Although inclusive and applicable to a broad set of use cases, this definition does not identify specific dimensions of quality data. The International Organization for Standardization (ISO) defines quality as “the totality of the characteristics of an entity that bear on its ability to satisfy stated and implied needs” [16, p. 301]. Researchers went further to identify and assess over forty quality dimensions prevalent in publications from 1985 to 2009. The most referenced dimensions of data quality were accuracy, completeness, consistency, and currency. Figure 2 illustrates that accuracy, completeness, and consistency relate to independent variables, while currency relates to dependent variables. Accuracy is the extent to which data is correct, reliable, and certified [17, p. 14]. Completeness is the “degree to which values are present in a data collection” [18, p. 7]. Consistency is “the extent to which data is presented in the same format and compatible with previous data” [17, p. 14]. Currency is “the degree to which a datum is up-to-date” [18, p. 8].

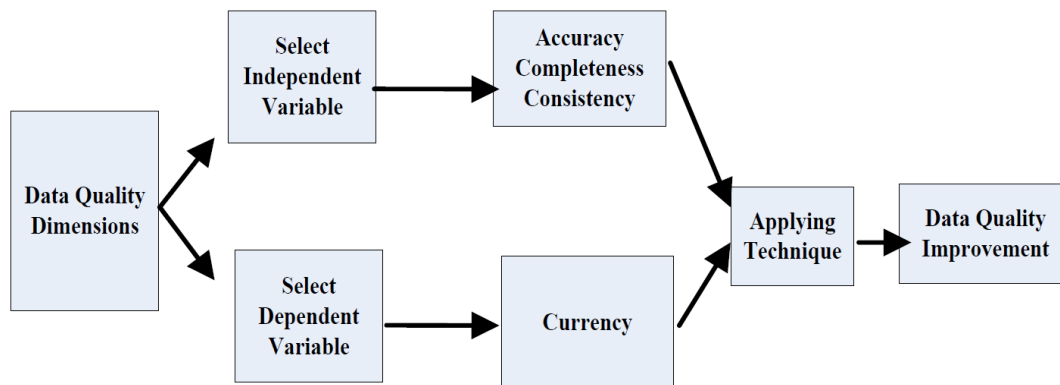


Figure 2: Data Quality Framework [16, p. 304]

These four dimensions (accuracy, completeness, consistency, and currency) are generally accepted to be critical to establishing data quality. Statistical techniques have been developed to assess some of these dimensions of quality. Accuracy is determined by assessing the number of true records compared with the number of false records. Error rate is the proportion of incorrect records [19, p. 320]. Precision is “defined as the proportion of positive examples that are truly positive” and indicates how trustworthy a sample dataset is [19, p. 329]. Recall indicates “how complete the results are” [19, p. 329]. Finally, the F_1 score (F -measure or F -score) combines precision and recall into a single number using the harmonic mean [19, p. 331].

2.7.2. Record Linkage & Approximate String Matching

Maintenance data from the Services contain significant amounts of free text input by users. Unlike coded entries or selections made from a menu of pre-defined inputs, free text is prone to significant variation. Record linkage, or data matching, is a field of computer science in which the task is to identify, match, and merge records that correspond to the same entities in one or several databases. The challenge is the lack of common entity identifiers in the data sources to be matched. The challenge of matching records using attributes that contain only partially identifying information, due to partial availability of data, typographical variations present in the data, or changing information input over time [10, p. 11].

Record linkage is a complex process with several steps (shown in Figure 3):

1. Data pre-processing (data cleaning and standardization) is crucial for quality record linkage outcomes because most real-world data contain noisy, incomplete, and inconsistent data. This step includes filling in missing data, removing unwanted values, transforming data into well-defined and consistent forms, and resolving inconsistencies in data representations and encodings.
2. Indexing is aimed at reducing the number of comparisons that need to be conducted between records. Record pairs that are as possible that are unlikely to correspond to matches are removed. Only pairs that are potentially matching, are brought together to be compared in detail in the next step.
3. Comparison requires using appropriate similarity functions, depending on the type of data that is compared. Approximate string matching functions are typically used for text record pairs, while different comparison functions are employed for dates and numeric values.
4. Classification involves assessing the numeric similarity values of all compared attributes. The similarity vectors of the compared candidate record pairs are given to a decision model which will classify record pairs into matches (where it is assumed that the two records in the pair correspond to the same entity), non-matches (where it is assumed that the two records in the pair correspond to different entities), and possible matches (where the classification model cannot make a clear decision). If record pairs are classified as possible matches, a clerical review process is required where these pairs are manually compared and classified into matches or non-matches.
5. Evaluation requires assessing the quality of the record linkages to a source of ground truth. In practice, evaluation is difficult because having access to a baseline dataset for comparison is uncommon [20, pp. 948-949].

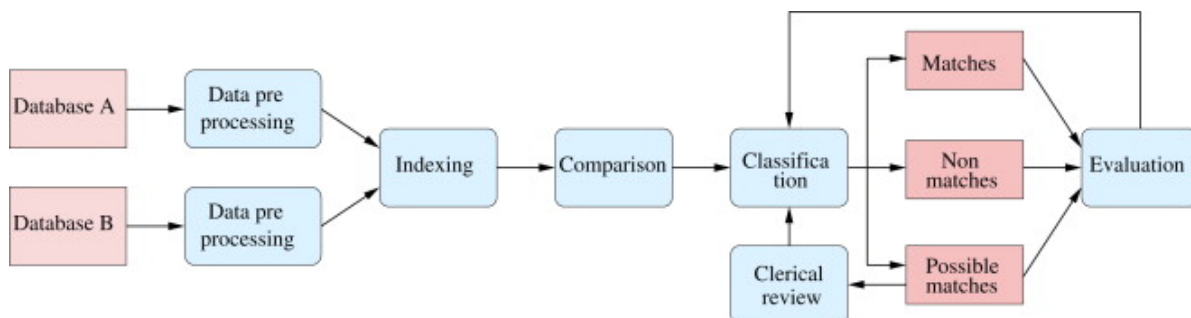


Figure 3: General Linkage Process [20, p. 948]

These approaches require an approximate comparison function. To compare H-60 and HMMWV data from the Services. String matching is a technique that would identify the exact matches of text from different data fields. More appropriate for free text field analysis, such as the maintenance data for military platforms, is approximate string matching. Approximate string matching allows errors within a text string and recognizes patterns within compared text fields [21, p. 31]. This technique then identifies the “edit distance” between the compared strings, where edits are defined as insertions, deletions, substitutions, or transpositions. Fewer edits represent a closer match, and therefore a better score [21, p. 37].

To assess the consistency of the maintenance data, a record pair comparison method that is suited to the data is required. In the case of the maintenance data from the Services, text comparisons between defined coded fields can be compared with free text narratives input by the maintainer. Depending on the coded field, the narrative should contain some related term(s), which can then be compared using approximate string matching. Several approximate string matching algorithms exist and were considered for this study. These include the Jaro-Winkler, Levenshtein edit distance, Editext, Longest Common Substring, Bag Distance, Ratcliff-Obershelp, Monge-Elkin, and Jaccard algorithms. The Jaro-Winkler edit distance algorithm, paired with text pre-processing, was best suited to compare unique words and phrases from the maintenance data sets. Text pre-processing ensures that data used for comparison are “in the same format, structure, and content. This includes removing unwanted characters and words, expanding abbreviations, correcting misspellings, and splitting or combining attributes” [10, p. 11].

The Jaro-Winkler Algorithm is based on the Jaro record linkage algorithm, which was developed as a matching technique for records from the US Census in the 1970s. The Jaro Algorithm has a simple formula to calculate the similarity of two strings based on respective length, number of common characters, and number of character transpositions [22, p. 87]. Jaro’s formula to compute similarity score is:

$$sim_j = \begin{cases} 0 & \text{if } m = 0 \\ \frac{1}{3} \left(\frac{m}{|s|} + \frac{m}{|t|} + \frac{m - \tau}{m} \right) & \text{otherwise} \end{cases}$$

$$match\ distance\ test = \lfloor \max(|s|, |t|) / 2 \rfloor - 1$$

where $|s|$ and $|t|$ are the lengths of the two strings being compared, m is the number of matching characters (number of characters that are identical and no further apart than the match distance, and τ is the number of character transpositions (half the number of matching characters in a different sequence order). If no characters match between strings s and t , then $m = 0$, and their Jaro similarity (sim_j) is 0. On the other hand, if every character matches and their character sequences are the same, $|s| = |t|$ and $\tau = 0$, then $sim_j = 1$ [23, pp. 2469-2470].

William Winkler found evidence that the “the probability of keypunch errors increased as the character position in a string moved from the left to the right” [24, p. 11]. As such, he developed the Jaro-Winkler Algorithm to calculate the similarity of two strings based on respective lengths, number of common characters, and number of character transpositions, *while rewarding matching characters at the beginning of a string*. The Jaro-Winkler formula builds upon Jaro’s formula and is:

$$sim_{jw} = \begin{cases} sim_j & \text{if } sim_j < b_t \\ sim_j + (l * p * (1 - sim_j)) & \text{otherwise} \end{cases}$$

where p and b_t are typically set at 0.1 and 0.7. p is a scaling factor to adjust upwards for common prefixes and b_t is a predetermined boost threshold, and l is the length of s and t ’s common prefix ($l \leq 4$). If $sim_j < b_t$ then $sim_{jw} = sim_j$. Otherwise, the Jaro-Winkler similarity score (sim_{jw}) is augmented by the product of the common prefix between the two strings ($l \leq 4$), a scaling factor (p), and the unrealized similarity ($1 - sim_j$). As with the Jaro Algorithm, Jaro-Winkler similarity scores range from 0 to 1, where 0 represents no characters match between strings and 1 represents all characters match between strings [23, p. 2470].



2.8. Hypotheses

Consistency and completeness are the main components assessed for this study. Each hypothesis below is derived from a study objective so that each objective could be thoroughly investigated and evaluated.

2.8.1. Hypothesis 1 – Navy H-60

Hypothesis 1 is linked to Objective 1 of this study, which is to assess the data quality for U.S. Navy and U.S. Air Force H-60 rotary wing aircraft maintenance data. Hypothesis 1 focuses on Navy H-60 data (both organizational and depot-level), and considers the object receiving maintenance, maintenance actions taken, malfunction noted, and corrosion tagging.

Hypothesis 1a (object): If Navy H-60 data maintenance records are consistent, then free text maintenance data fields (discrepancy and corrective action) should include the specific system, subsystem, or component code. (*Free text fields vs. Work Unit Code (WUC)*)

Hypothesis 1b (action): If Navy H-60 data maintenance records are consistent, then the free text corrective action data field should include keywords from the action taken code and/or work unit code. (*Corrective action vs. AT Code / WUC*)

Hypothesis 1c (malfunction): If Navy H-60 data maintenance records are consistent, then free text descriptions should include keywords from the malfunction code. (*Free text fields vs. Malfunction Codes (MALCODES)*)

Hypothesis 1d (corrosion-related): If Navy H-60 data maintenance records are consistent, then free text descriptions should include corrosion keywords when a record has a corrosion-related codes assigned to it. (*Free text fields vs. Malfunction Codes (MALCODES)*)

2.8.2. Hypothesis 2 – Air Force H-60

Hypothesis 2 is linked to Objective 1 of this study, which is to assess the data quality for U.S. Navy and U.S. Air Force H-60 rotary wing aircraft maintenance data. Hypothesis 2 focuses on Air Force H-60 data (both organizational and depot-level), and considers the object receiving maintenance, maintenance actions taken, malfunction noted, and corrosion tagging.

Hypothesis 2a (object): If Air Force H-60 data maintenance records are consistent, then free text maintenance data fields (maintenance action discrepancy and maintenance corrective action) should include the specific system, subsystem, or component code. (*Free text fields vs. Work Unit Code (WUC)*)

Hypothesis 2b (action): If Air Force H-60 data maintenance records are consistent, then the free text maintenance action data field should include keywords from the action taken code and/or work unit code. (*Maintenance corrective action vs. AT Code / WUC*)

Hypothesis 2c (malfunction): If Air Force H-60 data maintenance records are consistent, then free text descriptions should include keywords from the malfunction code. (*Free text fields vs. HOW_MAL code*)

Hypothesis 2d (corrosion-related): If Air Force H-60 data maintenance records are consistent, then free text descriptions should include corrosion keywords when a record has a corrosion-related codes assigned to it. (*Free text fields vs. HOW_MAL code*)

2.8.3. Hypothesis 3 – Army HMMWV

Hypothesis 3 is linked to Objective 3 of this study, which is to assess the data quality for the HMMWV across Services. The U.S. Army and U.S. Marine Corps are the only branches of Service where HMMWVs are reportable items. Therefore, data analysis was focused on Army and Marine Corps HMMWVs (both up-armored and non-up-armored). Unlike the Marine Corps HMMWV data, for the Army, enough data fields are available in the data set to compare the object receiving maintenance, maintenance action taken, malfunction noted, and corrosion tagging.

Hypothesis 3a (object): If Army HMMWV data maintenance records are consistent, then free text maintenance data fields (fault description, correction narrative, and remarks) should include the specific system, subsystem, or component code. (*Free text fields vs. part nomenclature*)

Hypothesis 3b (action): If Army HMMWV data maintenance records are consistent, then the free text correction narrative data fields should include keywords from the action taken code. (*Correction narrative vs. action description*)

Hypothesis 3c (malfunction): If Army HMMWV data maintenance records are consistent, then free text descriptions should include keywords from the failure code. (*Free text fields vs. failure code*)

Hypothesis 3d (corrosion-related): If Army HMMWV data maintenance records are consistent, then free text descriptions should include corrosion keywords when a record has a corrosion-related codes assigned to it. (*Free text fields vs. failure code*)

2.8.4. Hypothesis 4 – Marine Corps HMMWV

Hypothesis 4 is linked to Objective 3 of this study, which is to assess the data quality for the HMMWV across Services. For the Marine Corps, there are limited data fields available, so the focus is on determining if the free text field aligns with the defect code(s).

Hypothesis 4a (malfunction): If Marine Corps HMMWV data maintenance records are consistent, then free text descriptions should include keywords from the defect code. (*Free text fields vs. defect code(s)*)

Hypothesis 4b (corrosion-related): If Marine Corps HMMWV data maintenance records are consistent, then free text descriptions should include corrosion keywords when a record has a corrosion-related codes assigned to it. (*Problem summary vs. defect code(s)*)

2.9. Study Limitations

Limitations are circumstances limiting the study team's ability to fully meet the study objectives or fully investigate the problem. The outcomes from this study are limited based on two main factors that impacted the research. The first issue relates to the quality of the raw maintenance data received from the Services,



and the second issue relates to the COVID-19 environment that impacted the ability to conduct research and interviews with subject matter experts across maintenance locations during the duration of this study.

Maintenance data, especially free text entry in maintenance records, is highly variable. Descriptions may lack adequate detail, contain shorthand and slang, and include erroneous information. This is a significant limitation to truly assessing all aspects of data quality. As such, the analyst focused analytical efforts on data consistency and data completeness, as those can be assessed without making assumptions as to what the user intended with free text entries. Below provides an example of actual of a free text entry from a single Air Force HH-60 maintenance record:

Maintenance Action Discrepancy:

PFA P325 REQ'S INSTALL AFTER EOT CK (SFD 20161029 PG 2 I 2)

Maintenance Action Corrective Action:

REINSTALLED PFA P325 IAW TO 1H-60(H)G-2-9 PARA 10.9.1 STEP AX TOOL AND FOD CK C/W EOT CK GOOD SFD 20170314 PG 8 IT 3

This example illustrates some of the variability and challenges with free text entries. The “Maintenance Action Discrepancy” contains only two English words out of nine total terms. The “Maintenance Action Corrective Action” contains eight English words out of twenty terms. This proportion of non-words is not uncommon when assessing free text maintenance records. To mitigate some of the issues with irregular words, the methodology allows for single-word pairwise comparisons. After removing common “stop words,” this pairwise comparison is useful to identify relevant common or similar terms between strings.

The second limitation that impacted this study was the COVID-19 pandemic. For the significant majority of this research, all government travel was prohibited, and most personnel were working remotely from home. These conditions limited the ability to travel to maintenance locations and directly observe maintenance data collection processes. It also severely inhibited the opportunity to meet and connect with maintenance subject matter experts as well as network to gain access to the most appropriate data sets. While CPO and the CCPEs were supportive in trying to make the right connections for the analyst, there was not the ability to find an appropriate “ground truth” data set. Having a verified data set was required to make meaningful assessments on data accuracy. Without a “ground truth” data set that confirms the actual maintenance conditions that existed, data accuracy cannot be fairly assessed. As such, the hypotheses and study focus on data consistency and data completeness.



Section 3

Approach and Methodology

3.1. General Approach

The general approach for this study was divided into two primary efforts in support of the four research objectives.

The first line of effort focused on study Objectives 1 and 2, which was to assess the consistency and completeness of U.S. Navy and U.S. Air Force H-60 rotary wing aircraft maintenance data. Specifically, the analyst considered the consistency and completeness of data quality for the Services regarding the object receiving maintenance, the maintenance action taken, the maintenance malfunction, and whether a maintenance malfunction was corrosion related. The research approach was the application of record linkage and approximate string-matching techniques to identify the consistency of maintenance entries by a user or maintainer. After assessing data consistency and completeness, the second objective was to identify best practices from across the Services and generate recommendations regarding corrosion-related maintenance data collection. These best practices and recommendations are based on the data quality assessments as well as interviews and research into maintenance data collection processes. Objective 2 is discussed in Section 4.2 after presenting the findings from the Navy and Air Force H-60 maintenance data quality analysis.

The second line of effort focused on study Objectives 3 and 4 as stated in Section 1.2. Objective 3 assesses data quality for the HMMWV to determine scalability of the research approach used on H-60 rotary wing aircraft to a cross-Service ground system. Similar to the H-60 data quality analysis, the study assesses quality for the object receiving maintenance, the maintenance action taken, the maintenance malfunction, and whether a maintenance malfunction was corrosion related. After assessing the HMMWV, Objective 4 put forth recommendations for additional ground systems for future analysis. Objective 4 is discussed in Section 4.4 after presenting the findings from the HMMWV maintenance data quality analysis.

The general methodology for this study is shown in Figure 4. Phase 1 of this effort focused on researching how maintenance data is collected across Services in the DoD. This knowledge helps to identify data sources and how the Services differ in their maintenance procedures. Phase 2 is the data quality analysis and assessment of H-60 maintenance data for the Navy and Air Force. Phase 1 and Phase 2 both help to inform Phase 3, which is the identification of best practices for maintenance data collection across the DoD. Phase 4 uses a similar approach to Phase 2 to assess ground vehicle data for a cross-Service platform. The study assesses the HMMWV (both up-armored and non-armored) as the ground vehicle. Phase 4 demonstrates how the methodology may be scaled to other platforms beyond rotary wing aviation systems. Phase 5 recommends additional ground systems to which the methodology can be extended going forward. These phases generally align with the study objectives and the overall methodology is clearly related.

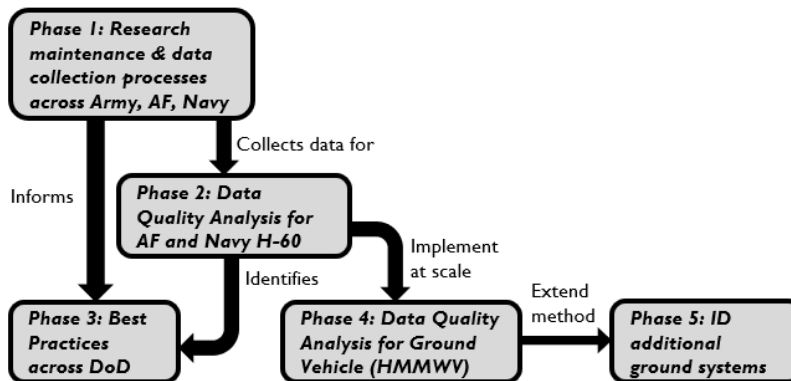


Figure 4: Study Methodology

3.2. Objective 1: Maintenance Data Quality for Navy & Air Force H-60

The following sections discuss the data sources used and research procedures that was applied for the data. The findings from this research approach are discussed in Section 4.1.

3.2.1. Data Sample

Collecting H-60 maintenance data for Navy and Air Force required communication, coordination, and approval from several organizations. Targeted efforts were made to collect organizational-level and depot-level data from the Navy and Air Force. This section details the data sources and summarizes the data obtained from each source.

The Navy Program Management Activity (PMA) 299 provides full-spectrum support for the Navy's H-60 fleet, and subject matter experts from PMA-299 provided organizational and depot-level data in support of this study [25]. The maintenance data was isolated to calendar year 2017 to align with the data obtained by MAJ John Case in his FY20 CPO study that assessed maintenance data for the Army H-60 system.

Organizational data from PMA-299 included MH-60 data from OOMA from January through December 2017. The dataset included a total of 147,795 maintenance records capturing discrepancies and actions taken for MH-60s across various naval bases. Table 1 contains a partial sample from the dataset, illustrating some of the key data fields available. The complete dataset contained 29 data fields containing both coded fields and free text fields. Critical coded fields for this study include: WUC, AT Code, and MALCODES. Free text fields necessary for this study include: System Reason Description, Discrepancy, and Corrective Action.

**Table 1: Navy H-60 Organizational Level Maintenance Data Sample**

WUC	CompDateTime	Buno	SystemReasonDescription	MalfunctionCode	ActionTakenCode	Discrepancy	CorrectiveAction
111S140	20-Mar-17	168580	CRTCH PNL SCRW CCDD:17106	170	Z	CROTCH PANEL SCREW CORRODED. CCDD: 17106	REMOVED CORROSION THEN CLEANED, INSPECTED, & TREATED CROTCH PANEL SCREWS IAW 01-1A-509 SERIES. AREA SECURE AND FOD CHECK COMPLETE.
1215C40	21-Mar-17	166294	CAB SEAT BAR CCDD:29FEB17	170	C	CABIN SEAT BAR HW CORRODED	REMOVED AND REPLACED CABIN SEAT BAR HW IAW 110300 WP 14 AREA SECURE AND FOD CHECK COMPLETE ATAF
71Y9100	15-Feb-17	166294	DFG ANTENNA	170	R	DFG ANTENNA IS CORRODED BEYOND REPAIR	REMOVED AND REPLACED DFG ANTENNA IAW 710320 AREA CLEAN AND FOD FREE PEMA 6518
57R1G10	2-Mar-17	166294	LOWER STAB ACTUATOR	170	B	LOWER STAB ACTUATOR CLEVIS AND ATTACHING HARDWARE ARE CORRODED BEYOND REPAIR	REMOVED CORROSION FROM LOWER CLEVIS CLEANED TREATED AND PRESERVED IAW 011A509 REMOVED AND REPLACED ATTACHING HARDWARE IAW 110300 AREA CLEAN AND FOD FREE PEMA 6518

Depot data from PMA-299 included MH-60 data from ADCS from January through December 2017. The dataset included a total of 19,838 maintenance records capturing discrepancies and actions taken for MH-60s from the Navy's three major depot maintenance locations (Fleet Readiness Center South East in Jacksonville, FL, Fleet Readiness Center Mid-Atlantic in Norfolk, VA, and Fleet Readiness Center South West in North Island, CA). Table 2 contains a partial sample from the dataset, illustrating some of the key data fields available. The complete dataset contains 14 data fields containing both coded fields and free text fields. Critical coded fields include: WUC and MALCODES. The Navy Depot-Level data does not include Action Taken codes, so Hypothesis 1b will not be assessed for Navy depot-level maintenance in this study. The only free text field available in the dataset is the Discrepancy Description field, which will allow for assessing Hypothesis 1a, 1c, and 1d.

Table 2: Navy H-60 Depot Level Maintenance Data Sample

ClassDescription	Type Corrosion	Aircraft Zone	DiscrepancyDescription	Level Maintenance	MAL Code	WUC	Specification Code
Major	Pitting	01	PILOT'S COCKPIT DECK: FS 246, BL 38R-43R, WL 216-242.5 REMOVE CORROSION ON ATTACH BRACKET AND ON/UNDER A/NUTS AS MARKED. CERT RESTORE SURFACE FINISH AND ENSURE FO REMOVAL	DEPOT	170	111R2	SS1I
Major	Exfoliation	01	STATION 247, CO-PILOT DECK: FS 247, BL 38L, WL 216 REMOVE AND TREAT CORROSION ON BRACKET AND A/NUTS AS MARKED. CERT RESTORE SURFACE FINISH AND ENSURE FO REMOVAL	DEPOT	170	11210	RS1B
Major	Surface	01	NOSE EXTERIOR: á FS 185, BL 11.5L-30L, WL 205 á REMOVE CORROSION ON/AROUND A/NUTS AND ATTACH FLANGE AS MARKED. RESTORE SURFACE FINISH AND ENSURE FO REMOVAL	DEPOT	170	111R2	SS1I
Major	Pitting	02	KIT J444 FAIRING, FUSELAGE: á P/N 70215-02319-041 á REPLACE CORRODED A/NUTS AS MARKED/REQUIRED. á CERT á RESTORE SURFACE FINISH. á CERT á ENSURE FOD REMOVAL. á CERT á	DEPOT	170	11250	

After collecting Navy O-level and D-level maintenance data, the specific coded fields need to be translated into text fields to allow for record linkage and approximate string matching techniques to be employed. The primary source used for obtaining the coded field descriptions was the NAMP, codified in COMNAVAIRFORINSTR 4790.2C. NAMP Appendix E specifies Maintenance Documentation Codes. AT Codes and MALCODES are provided, with 29 unique AT Codes and 322 MALCODES [12]. To obtain the WUC descriptions, the analyst gained access to the Decision Knowledge Programming for Logistics Analysis and Technical Evaluation (DECKPLATE). DECKPLATE is a data warehouse platform used by NAVAIR to streamline and centralize fleet management [26]. DECKPLATE has 7,399 unique WUCs that apply to the H-60 platform.

To assess corrosion-related free text fields, an authoritative listing of corrosion terms is required. MADW has developed, in conjunction with each Service, a "corrosion dictionary." Table 3 illustrates a sample listing



from the corrosion dictionary, which contains 2,816 terms and phrases used to identify and predict if a record is likely corrosion related. The MADW Corrosion Dictionary is unique for each Service and function (e.g., Navy aviation vs. Marine Corps ground vehicles) This listing is incorporated to the assessment for Hypothesis 1d.

Table 3: MADW Navy/Marine Corps Aviation Corrosion Dictionary Sample

SearchTerm	ContainsOrLike	CorroKeyword	GrossOrder	Source	MaintOperationKeyword
"Aerodynamic Smoothing Compound**"	Contains	AERODYNAMIC SMOOTHING COMPOUND	1	Corro	REPAIR
"Critical pitting potential**"	Contains	CRITICAL PITTING POTENTIAL	2	Corro	REPAIR
"Oxygen concentration cell**"	Contains	OXYGEN CONCENTRATION CELL	3	Corro	REPAIR
"Gallionella ferruginea**"	Contains	GALLIONELLA FERRUGINEA	4	Corro	REPAIR
"Open-circuit potential**"	Contains	OPEN-CIRCUIT POTENTIAL	5	Corro	REPLACE

The Air Force Corrosion Prevention and Control Office (AFCPCO) provided maintenance data for the HH-60G platform for organizational level and depot level. AFCPCO supports the Air Force Life Cycle Management Center (AFLCMC), which is one of six centers within Air Force Materiel Command (AFMC). AFCPCO is located at Robins Air Force Base in Georgia and is the Air Force's "single focus point for all system corrosion guidance and policy" [27]. AFCPCO "provide[s] world-class aerospace corrosion expertise, guidance and training to maximize mission effectiveness, increase safety and reduce cost through improved materials and processes" [27]. AFCPCO subject matter experts provided organizational and depot-level data in support of this study. The maintenance data was isolated to calendar year 2017 to align with the data obtained by MAJ John Case in his FY20 CPO study that assessed maintenance data for the Army H-60 system.

Organizational maintenance data from AFCPCO included HH-60G data from REMIS from January through December 2017. The dataset included a total of 50,760 maintenance records capturing discrepancies and actions taken for HH-60G vehicles across various Air Force bases. Table 4 contains a partial sample from the dataset, illustrating some of the key data fields available. The complete dataset contained 23 data fields containing both coded fields and free text fields. Critical coded fields for this study include: WUC, AT Code, and MALCODES. Free text fields necessary for this study include: Work Order Discrepancy, Maintenance Action Discrepancy, and Maintenance Corrective Action.

Table 4: Air Force H-60 Organizational Level Maintenance Data Sample

EQUIP_DESIG_ID	WO_Discrep	WUC	AT_Codes	MALCODES	Maint_Action_Discrep	Maint_Action_Corr_Action
HH060G	MGB FWD FAIRING (DOG HOUSE) HAND HOLDS DELAMINATED DISBANDED AND CRACKED	1131B	F	846	MGB FWD FAIRING (DOG HOUSE) HAND HOLDS DELAMINATED DISBANDED AND CRACKED	MGB FWD FAIRING HAND HOLDS REPAIRED AS REQUIRED IAW 1H-60(H)G-2-2 FIG. 1-37
HH060G	L/H SLIDING PYLON COVER ACFT MOUNTED , MIDDLE PIN RECEIVER WORN	1115H	Z	170	L/H SLIDING PYLON COVER ACFT MOUNTED , MIDDLE PIN RECEIVER WORN	PIN RECEIVER CLEANED AND TREATED.
HH060G	CABIN FURNISHINGS REMOVED FOR 600 PE IAW 1H60(H)G-6WC-2	11200	X	799	CABIN SOUNDPROOFING REQUIRE RIVETS FOR INSTALLATION	RIVETS REPLACED AS REQUIRED
HH060G	NOSE SHELF REQ'S CORBAN	11114	Z	865	NOSE SHELF REQ'S CORBAN	APPLIED CORBAN
HH060G	T/G/B FORWARD COWLING REQUIRES SEAL ON OUTSIDE EDGE NEXT TO TAIL ROTOR SLIP RING ASSY.	1151B	G	730	T/G/B FORWARD COWLING REQUIRES SEAL ON OUTSIDE EDGE NEXT TO TAIL ROTOR SLIP RING ASSY.	REPLACED T/G/B FORWARD FAIRING SEAL



Depot maintenance data for Air Force HH-60G was attained by collaborating with AFCPCO and the Mercer Engineering Research Center (MERC). MERC provides expert engineering support to Robins Air Force Base in several engineering fields to include corrosion and obsolescence [28]. The depot maintenance data provided HH-60G data from REMIS from January through December 2017. The dataset included a total of 29,484 maintenance records for HH-60G vehicles from the main locations where Air Force depot maintenance is conducted (Jacksonville, FL, Lexington, KY, Corpus Christie, TX, and Korea). Table 5 contains a partial sample from the dataset, illustrating some of the key data fields available. The available depot data format and depth varied depending on location, ranging from 13 to 32 available data fields containing both coded fields and free text fields. Critical coded fields for this study include: WUC, AT Code, and MALCODES. Free text fields necessary for this study include: Work Order Discrepancy, Maintenance Action Discrepancy, and Maintenance Corrective Action. Not every depot maintenance data file contained all coded fields not related free text fields; however, for this study, if the fields were present, they were assessed.

Table 5: Air Force H-60 Depot Level Maintenance Data Sample

Seq Num	Fault Date	Fault Id	Fault Desc	Instruction
7PJ04	9/20/2017	1874216	PRECHECKS REFUEL PROBE MOUNT REPAIR INSTALL (TCTO 1H-60-1049)	Perform a main and tail rotor rig check in accordance with T.O. 1H-60(H)G-2-5 and TO 1H-60-1049 Step 6.b. Record rig check data for comparison once the repair has been completed.
7PJ04	9/20/2017	1874217	REMOVE/INSTALL REFUEL PROBE MOUNT REPAIR INSTALL (TCTO 1H-60-1049)	Install floor skin panel I/A/W TO 1H-60(H)G-2-2.
7PJ04	9/23/2017	1874532	REPAIR DOOR ASSY, CARGO L/H	Repair as directed by PSA I/A/W TO 1H-60(H)G-2-2 para 3.5.
7PJ04	9/23/2017	1874533	PRIME/PAINT DOOR ASSY, CARGO L/H	Prime interior with Epoxy, MIL-PRF-23377. Prime all bare metal parts (handles, latches, hinges, etc.) as required.

As with Navy H-60 maintenance data, Air Force coded fields need to be translated into text to employ record linkage and approximate string matching techniques. The primary source used to match coded fields with their respective text descriptions was the *Air Force Technical Order (TO) 1H-60(H)G-06: Organizational Maintenance – Work Unit Code Manual*. TO 1H-60(H)G-06 contains AT Codes, MALCODES, and WUC Codes. There are with 33 unique AT Codes, 225 MALCODES, and 2,319 WUC Codes [29]. MADW has developed, in conjunction with the Air Force, a corrosion dictionary like the Navy's. Table 6 illustrates a sample listing from the corrosion dictionary, which contains 2,828 terms and phrases used to identify and predict if a record is likely corrosion related. This listing is incorporated to the assessment for Hypothesis 2d.

Table 6: MADW Air Force Corrosion Dictionary Sample

SearchTerm	ContainsOrLike	MaintObject	GrossOrder	Source	MaintOperationKeyword
"Semipermeable membrane**"	Contains	MEMBRANE	6	Corro	REPLACE
"Caustic embrittlement**"	Contains	CAUSTIC EMBRITTLEMENT	7	Corro	REPAIR
"Sulfate-reducing bact**"	Contains	SULFIDE	8	Corro	REPLACE
"surface active agent**"	Contains	SURFACE ACTIVE AGENT	9	Corro	REPAIR
"Breakdown potential**"	Contains	BREAKDOWN POTENTIAL	10	Corro	REPAIR



3.2.2. Research Procedures

For Objective 1, the goal was to identify how consistent the Navy and Air Force H-60 maintenance data was through comparing coded data fields to free text data fields. The method employed to compare, classify, and evaluate data consistency is the Jaro-Winkler Algorithm. A major benefit of the Jaro-Winkler Algorithm is that it can be applied to any two distinct text data fields, regardless of source. The hypotheses being tested (see Sections 2.8.1 and 2.8.2) evaluate data consistency by assessing the Jaro-Winkler Algorithm output scores. For this study, R was used as the programming language, as it allows for text analysis, statistical analysis, data visualization, and is available through open-source means.

The record linkage process (see Section 2.7.2) was followed to explore the first objective. Data pre-processing was completed to make the text uniformly lowercase, remove unwanted characters and stop words, remove punctuation, expand abbreviations and splitting or combining attributes. Stop words refer to common words that are typically considered inconsequential for text analysis. Examples of stop words include: *'the,' 'to,' and 'and'* [30, p. 77]. Correcting misspellings was not performed, because data consistency requires assessing the data as-is, not correcting terms and removing important information from the data. Stemming was then completed on the text fields. Stemming refers to “removing suffixes and reducing a word to some base form such that all different variants of that word can be represented by the same form (e.g., “car” and “cars” are both reduced to “car”) [31, p. 45] [32, pp. 52-53]. To conduct meaningful comparisons, text strings from the coded and free text fields were tokenized. A token is a meaningful unit of text, such as a word or group of words, that can be used for analysis. “Tokenization is the process of splitting text into tokens” [33, pp. 1, 45]. An R package called “SnowballC” provided the ability to pre-process the text data in a standardized fashion and accounts for the adjustments as discussed [34, p. 242].

After conducting data pre-processing, a record linkage procedure was developed to allow for comparison between the coded data field descriptions and the free text fields. Figure 5 depicts the record linkage procedure used to assess the token comparisons conducted between the coded and free text data fields. Each coded data description was tokenized as were the respective free text data fields. Pairwise testing was performed using the Jaro-Winkler Algorithm for each “coded field token” to each “free text field token.” The R packages “data.table,” “stringr,” and “stringdist” provided the necessary functions to develop all possible combinations of token pairs for a maintenance record, conduct a Jaro-Winkler comparison, and output Jaro-Winkler similarity scores.

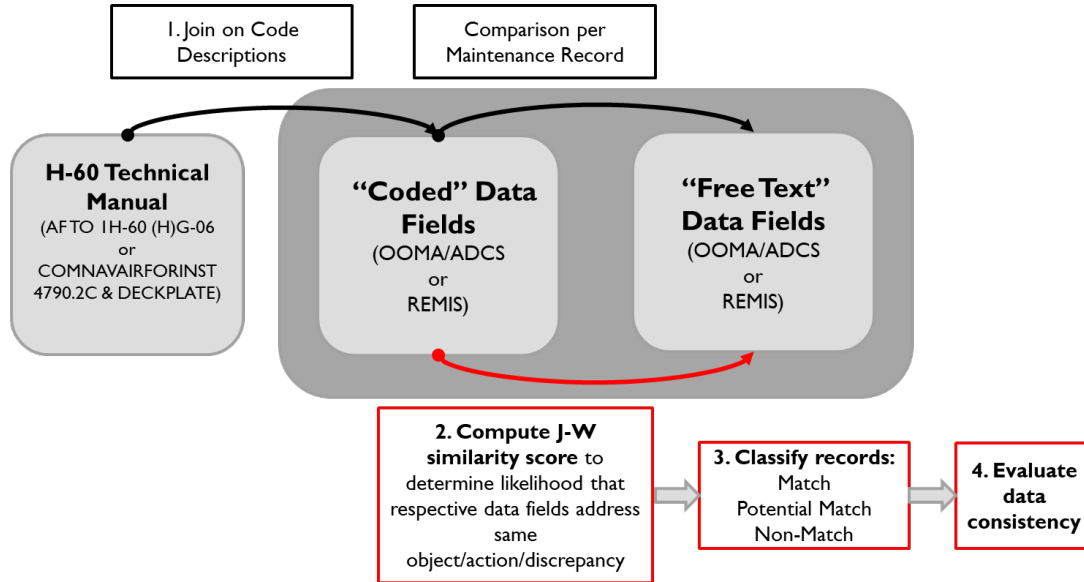


Figure 5: Record Linkage Procedure for H-60 Maintenance Data

As discussed, the Jaro-Winkler Algorithm was selected to conduct approximate string-matching for text data (tokens) for this study. The Jaro-Winkler Algorithm calculates the similarity of two tokens based on their respective lengths, number of common characters, and number of character transpositions, *while rewarding matching characters at the beginning of a token*. The Jaro-Winkler formula builds upon the Jaro Algorithm and is:

$$sim_j = \begin{cases} 0 & \text{if } m = 0 \\ \frac{1}{3} \left(\frac{m}{|s|} + \frac{m}{|t|} + \frac{m - \tau}{m} \right) & \text{otherwise} \end{cases}$$

$$match\ distance\ test = \lfloor \max(|s|, |t|) / 2 \rfloor - 1$$

$$sim_{jw} = \begin{cases} sim_j & \text{if } sim_j < b_t \\ sim_j + (l * p * (1 - sim_j)) & \text{otherwise} \end{cases}$$

where $|s|$ and $|t|$ are the lengths of the two token being compared, m is the number of matching characters (number of characters that are identical and no further apart than the match distance, and τ is the number of character transpositions (half the number of matching characters in a different sequence order). p is a scaling factor to adjust upwards for common prefixes and b_t is a predetermined boost threshold, and l is the length of s and t 's common prefix ($l \leq 4$). Jaro-Winkler similarity scores range from 0 to 1, where 0 represents

no characters match between strings and 1 represents all characters match between strings [23, pp. 2469-2470]. The highest Jaro-Winkler similarity score was retained per maintenance record (row). The justification is that after data pre-processing, the highest Jaro-Winkler similarity score for each records represents the greatest match likelihood between the fields. Since coded field descriptions are relatively short (typically 5 words or less), it is unlikely that extraneous words are assessed with this methodology, especially since stop words were removed during data pre-processing.

Record linkage requires the classification of records once they have been scored. The traditional method for record linkage is to have three classification bins: match, non-match, or potential match [20, p. 949]. This aligns well with the Jaro-Winkler Algorithm, as the similarity scores can be sorted into bins once the bins are defined. For this study, the analyst ran the Jaro-Winkler Algorithm over all available datasets and determined classification thresholds for match, non-match, and potential. Threshold-based classification is the most common method used in record linkage and is based on clerical review of comparison scores [20, p. 951]. After reviewing 100 records per dataset, thresholds were established for the classification bins. Table 7 shows the thresholds for each classification category.

Table 7: Classification Categories for J-W Similarity Scores

Classification Category	Similarity Score Interval
Match	[0.92 , 1.0]
Potential Match	[0.79 , 0.92)
Non-Match	[0 , 0.79)

After classifying records as match, non-match, or potential match, the final evaluation step was to determine the overall consistency match score. After reviewing 100 records in the 'potential match' bin, approximately 70% of 'potential matches' were deemed to be matches when comparing the stemmed coded data field description with the free text field. This allows for the overall calculation of estimated consistency match score. To calculate the estimated consistency match score, the following formula was used:

$$\text{Estimated Consistency Match Score} = \frac{\text{count of scores in } [0.92, 1.0] + p * \text{count of scores in } [0.79, 0.92)}{\text{total count of records}}$$

where p indicates the match proportion = (0.7)

This methodology was used for the Hypotheses 1a, 1b, 1c, 2a, 2b, and 2c. For each of these hypotheses, a distinct coded field exists that can be compared to a related free text field. For example, for the object receiving maintenance, the WUC description is directly compared to the combined corrective action and discrepancy fields. For the action taken-related hypotheses, the AT Code description is directly compared to the corrective action field. For the malfunction-related hypothesis, the MALCODE description is directly compared to the discrepancy field.

Hypotheses 1d and 2d (see Section 2.8.1 and 2.8.2) sought to address if a corrosion-related MALCODE contains corrosion-related free text when compared to respective MADW Corrosion Dictionary. To test these hypotheses, a modification to the previously-developed methodology was necessary. This modification allowed for free-text data fields to be compared to the MADW Corrosion Dictionary. Figure 6 illustrates the approach for record linkage while incorporating the relevant MADW Corrosion Dictionary. Like the previous methodology, step 1 was to join coded descriptions to coded fields within a dataset. Step 2 was to determine if a record has a corrosion-related MALCODE. The Navy and Air Force each have specific MALCODES that are corrosion-related, so those were used to split the datasets. The Navy has 49 corrosion-specific MALCODES, and the Air Force has 10 corrosion-related MALCODES (see Appendix D and Appendix E). Step 3 was to compare free text fields to the respective Service's MADW Corrosion Dictionary. The Jaro-Winkler Algorithm was again used to produce Jaro-Winkler similarity scores. After review, a threshold of 0.95 as assessed to split the match versus non-match classification bins. No potential match bin existed because of the need to build an error matrix (see Figure 7). After assigning Jaro-Winkler similarity scores, in step 4 the tokens were compared to MADW poison words (also called MADW exclusionary words). These are words and phrases that indicate the record is not corrosion related. Table 8 provides a sample of exclusionary words. If a record contained an exclusionary word, it was "untagged" as it was not a corrosion record. Finally, step 5 is the classification and evaluation of records based on the Jaro-Winkler similarity score and whether the MALCODE was corrosion-related. Records could fall into four classifications: Corrosion True Positive, Corrosion False Positive, Non-Corrosion True Negative, Non-Corrosion False Negative.

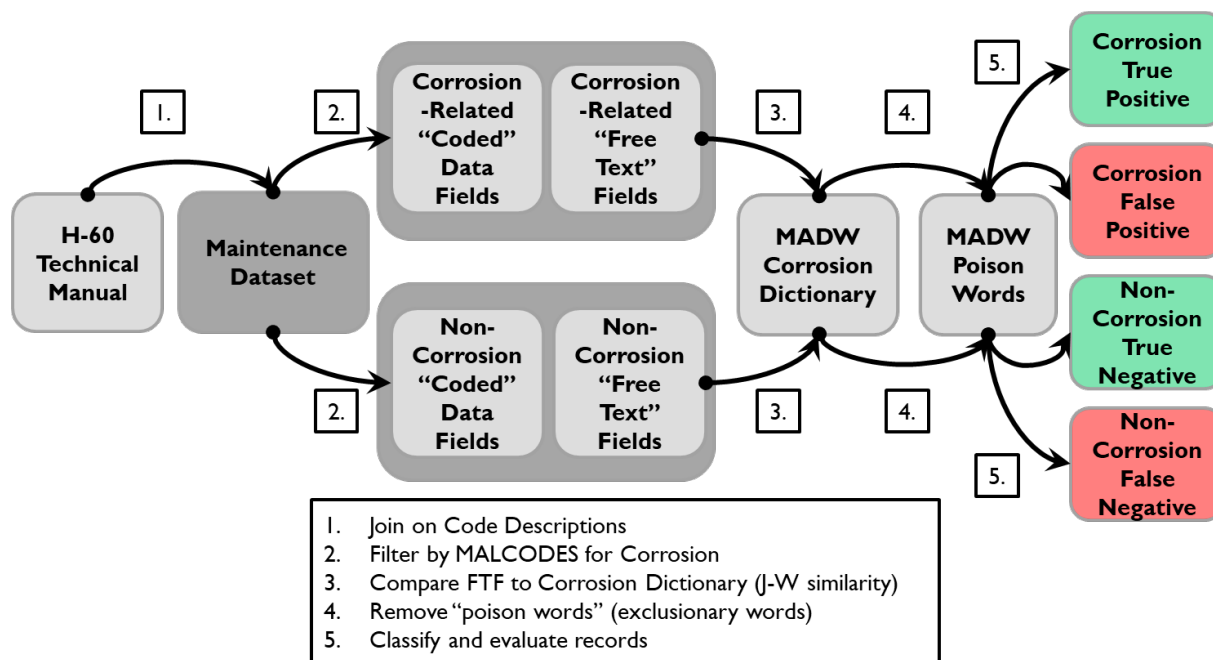


Figure 6: Record Linkage for Corrosion-Related Maintenance Records

Table 8: MADW Corrosion Poison/Exclusionary Words Sample

SearchTerm	ContainsOrLike	CorroKeyword
corrosion free	Contains	EXCLUDE-CORROSION FREE
not corrosion	Contains	EXCLUDE-NOT CORROSION
no corrosion	Contains	EXCLUDE-NO CORROSION
Internal failure*	Contains	EXCLUDE-INTERNAL FAILURE
Contamination*	Contains	EXCLUDE-CONTAMINATE

After following the record linkage methodology (see Figure 6), an error matrix (also called a confusion matrix) could be developed. An error matrix categorizes how well a model classifies values [19, p. 318]. An error matrix is typically used to assess a predictive model's ability to discern classes but can also be applied to assess whether two data fields are properly aligned [19]. Figure 7 is an error matrix as applied to the study. The error matrix asks two questions: (1) "Is the MALCODE corrosion-related?" and (2) "Does the Free Text Field contain Corrosion Keywords?". If the answer to question (1) is yes, then the record has a corrosion-positive code, else it has a corrosion-negative code. If the answer to question (2) is yes, then the free text is corrosion-positive, else the free text is corrosion-negative. After answering those questions, a record can be assessed as 'True Positive' (TP), 'False Positive' (FP), 'True Negative' (TN), or 'False Negative' (FN).

		Free Text Field contains MADW Corrosion Keywords?	
		Corrosion-Positive (A)	Corrosion-Negative (B)
Coded Data Field?	Corrosion-Positive (A)	True Positive (TP)	False Positive (FP)
	Corrosion-Negative (B)	False Negative (FN)	True Negative (TN)

Figure 7: Error Matrix for Corrosion-Related Maintenance Records

After developing the corrosion error matrices for the Navy and Air Force organizational and depot maintenance data, several performance measures can be calculated:

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$error\ rate = \frac{FP + FN}{TP + TN + FP + FN} = 1 - accuracy$$

$$precision = \frac{TP}{TP + FP}$$

$$recall = \frac{TP}{TP + FN}$$

$$F_1\ score = \frac{2 * precision * recall}{recall + precision} = \frac{2 * TP}{2 * TP + FP + FN}$$

These performance measures allow for comparison between Services and across maintenance levels. Section 4 addresses the performance measures as calculated from the datasets.

3.3. Objective 3: Maintenance Data Quality for HMMWV

The following sections discuss the data available, research procedures, and treatment of the data that was applied for the data. The findings from this research approach are discussed in Section 4.1.

3.3.1. Data Sample

Collecting HMMWV maintenance data for Marine Corps and Army required communication and coordination from several organizations. Targeted efforts were made to collect HMMWV maintenance data for both Up-Armored HMMWVs (UAH) and Non-Armored HMMWVs (NAH). However, when making data requests, it became apparent that selecting a specific UAH and NAH variant of the HMMWV would help target the data collection process. As such, two of the most common HMMWV platforms were selected, the M1151A1 (Truck, Utility: Expanded Capacity, Armament Carrier, IAP/Armor Ready) as a UAH variant and the M1097A2 (Truck, Utility: Heavy Variant, 4X4) as a NAH variant.

The Army's Combat Capabilities Development Command (CCDC) Data & Analysis Center (DAC) provided HMMWV organizational maintenance data from their Sample Data Collection (SDC) Program that includes M1151A1s and M1097A2s. The period of collection ran from October 2014 to September 2019 and collected on 1,667 unique serial numbers from select units at 21 locations. These locations included deployed exercises in Kuwait, Europe, and Korea; Combat Training Centers (CTCs) at Ft. Irwin and Ft. Polk; and various garrison locations across the United States and Germany. This data has been used to conduct analysis on HMMWV usage in non-combat conditions. The dataset contains 13,243 records of maintenance actions performed by serial number, along with the unit and location of each vehicle. Table 9 contains a partial sample from the dataset, illustrating some of the key data fields available. The complete dataset contains 29 data fields containing both coded fields and free text fields. Critical coded fields include: Part Nomenclature, Action Description, and Failure Description. Critical free text fields include: Fault Description, Correction Narrative, and Remarks. The SDC dataset was straightforward in the sense that all coded fields had their text descriptions already available, so there was no need to match coded fields to an external source. This dataset was ready for comparisons with minimal pre-processing required. The Marine



Corps & Army Ground Vehicle Corrosion Dictionary from MADW (sample shown in Table 11) was used to assess if free text fields contain corrosion-related terms and phrases.

Table 9: Army Organizational Level Maintenance Data Sample

Location	Model	Fault Date	Fault Description	Correction Narrative	action_description	Part Nomenclature	Failure Code	FailureDescription
FORT CARSON	M1097A2	2016-03-09	FRONT BLACKOUT LIGHT HOUSING WAS BROKEN	LIGHT BLACKOUT ORDERED AND REPLACED.	Replaced	LIGHT, BLACKOUT	099	OTHER/UNKNOWN /BROKEN
FORT HOOD	M1097A2	2016-08-31	ENGINE BELTS LOOSE	ADJUST BELT TENSION	Adjusted	BELTS, V, MATCHED SET	099	OTHER/UNKNOWN /BROKEN
FORT CARSON	M1097A2	2017-03-22	RADIATOR MISSING.	REPLACED RADIATOR.	Replaced	RADIATOR, ENGINE COOLANT	750	FAULT DUE TO MISSING COMPONENT(S)
FORT IRWIN	M1097A2	2017-07-11	FAN CLUTCH CLASS III LEAK.	REPLACED FAN CLUTCH.	Replaced	CLUTCH, FAN, ENGINE	381	FLUID LEAK
FORT CARSON	M1097A2	2017-10-03	MAINTENANCE PERFORMED A COOLANT TEST AND VEHICLES COOLANT FAILED.	FLUSHED AND REPLACED THE VEHICLES COOLANT.	Adjusted	RADIATOR, ENGINE COOLANT	790	ADJUSTMENT, INCLUDES OUT OF TOLERANCE

The Marine Corps' Materiel Management and Readiness Branch provided organizational-level maintenance data for UAH and NAH for calendar year 2019. The Materiel Management and Readiness Branch is part of the Logistics Plans, Policy, and Strategic Mobility Division, which reports to the Installations and Logistics Department of the Marine Corps. SMEs in the Materiel Management and Readiness Branch and the Deputy Commandant for Installation and Logistics supported the effort to retrieve and access organizational-level maintenance data for HMMWVs. The dataset included a total of 9,322 maintenance records capturing discrepancies M1151A1s and M1097A2s from across various the Marine Corps. Table 10 contains a partial sample from the dataset, illustrating some of the key data fields available. The complete dataset contained 44 data fields containing both coded fields and free text fields. The critical coded field available for this study is the DEF_CD_LST, which is a running list of defect codes for the specific maintenance entry. A given entry may contain one defect code or multiple. In this dataset, the largest number of defect codes for a single maintenance record was 15. Defect codes are distinguished with punctuation, either '|' or '.' to separate codes. The free text field in this dataset is "Problem Summary." A listing from the Materiel Management and Readiness Branch provides descriptions of all defect codes. Although there are limited coded fields and free text fields, the defect codes allow for identification of corrosion-related issues. Table 11 illustrates a sample listing from the Marine Corps & Army Ground Systems Corrosion Dictionary, which contains 4,189 terms and phrases curated by LMI to identify and predict if a record is likely corrosion-related. This listing is incorporated to the assessment for Hypothesis 4b.

Table 10: Marine Corps Organizational Level Maintenance Data Sample

PROBLEM_SUMMARY	DEF_CD_LST	SOURCE	NSN_IN_MAINTENANCE	SERIAL_NUMBER	MAINT_CATEGORY_CODE	DATE_CLOSED	MODEL_NUMBER
MASTER LIGHT SELECTOR IN OP	ELEC.RPLC	GCSS	2320015402038	628435	Operational - Minor	13-Jan-19	M1151A1
DEAD BATTERIES	ELEC. ELEC.INOP	GCSS	2320015402038	628538	Deadlined	3-Apr-19	M1151A1
P/S LAP BELT & POWER STEERING CAP REPLACEMENT	COMP.SAFDL	GCSS	2320015402038	645238	Operational - Minor	21-Feb-19	M1151A1
BIENNIAL PMCS FOR HMMWV SER # 601212	NMAJ.ASPM	GCSS	2320013808604	601212	Operational - Minor	4-Jan-19	M1097A2

Table 11: MADW Marine Corps & Army Ground Systems Corrosion Dictionary Sample

SearchTerm	ContainsOrLike	CorroKeyword	GrossOrder	Source	MaintOperationKeyword
"Mechanical biNDI/NDTng*"	Contains	MECHANICAL BINDI/NDTNG	16	Corro	REPLACE
"sodium bicarbonate*"	Contains	SODIUM BICARBONATE	17	Corro	DECONTAMINATE
"BLASTINGXXXREQUEST"	Contains	BLAST	18	Corro	BLAST
"Electrolytic cell*"	Contains	ELECTROLYTIC CELL	19	Corro	REPLACE
"incubation period*"	Contains	INCUBATION PERIOD	20	Corro	REPLACE

3.3.2. Research Procedures

In Objective 3, the goal was to identify how consistent the Marine Corps and Army HMMWV maintenance data was through comparing coded data fields to free text data fields. The method used is very similar to the approach employed for H-60 platforms. Figure 8 shows the procedure used, with the main adjustment being that the TMs and data fields align with Marine Corps and Army data sources. The primary purpose of this objective was to determine if the same research procedure is scalable to a ground system, so the intent was to keep the research procedures as similar as possible.

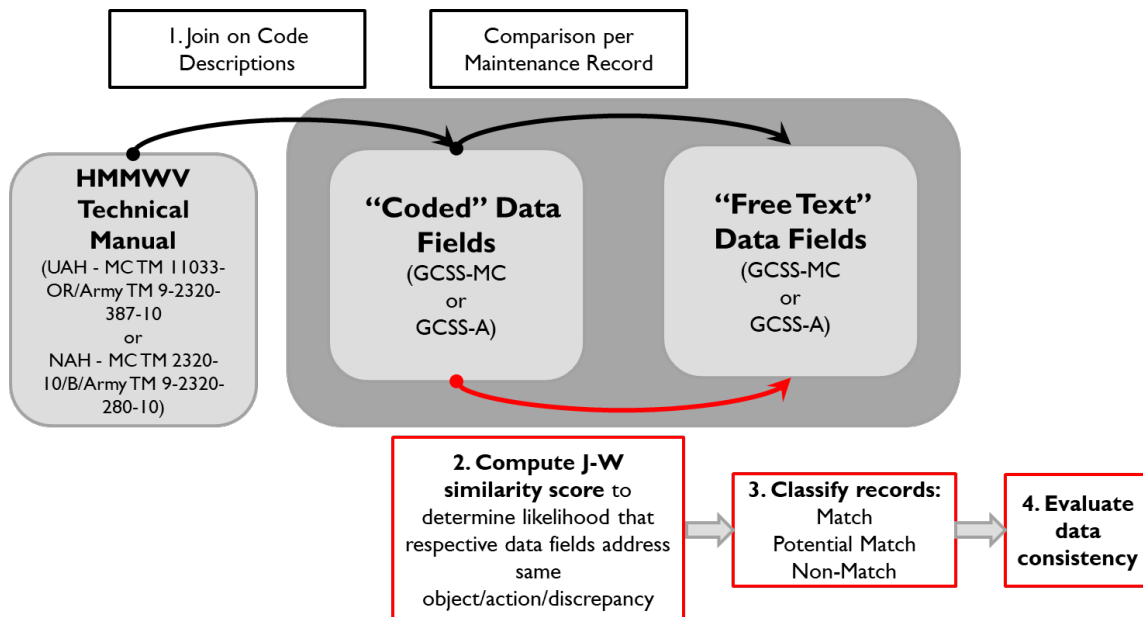


Figure 8: Record Linkage Procedure for HMMWV Maintenance Data

The available data for Marine Corps HMMWVs only allowed for limited analysis since there was one free text field and defect codes. Comparisons could be made to assess how consistency of the reported malfunction by comparing the defect code(s) to the problem summary free text field.

For the available Army HMMWV maintenance data, comparisons were made for the object receiving maintenance, the action taken, and the malfunction (Hypotheses 4a – 4c) by comparing coded fields to respective free text fields.



To generate the performance metrics and address Hypotheses 3b and 4d (*is the free text corrosion-related given a corrosion-related code?*), the same methodology shown in Figure 6 was used. This also allowed for the creation of an error matrix for Marine Corps HMMWV maintenance data and Army HMMWV maintenance data. The most significant difference for HMMWV MALCODES was the number of corrosion-related codes available. The Marine Corps data had four unique defect codes that related to corrosion (CORR, MOIST, SEAL, and PAINT). The Army data had one unique defect code that related to corrosion (Failure Code: 170). These represent significantly fewer MALCODES that relate to corrosion than the H-60 maintenance data had available.

3.4. Assumptions

Assumptions are statements that are taken as true in absence of facts, often to accommodate a limitation. This study relies on three critical assumptions. The first assumption relates to the detail included in the free text data. The second and third assumptions relate to the similarity score selection scoring methodology.

The first assumption made for this study is that free text data “provide[s] a narrative description that completely describes the problem. [Maintainers] provide as much detail as possible in failure analysis and help speed repairs... The corrective action narrative will contain a ‘free text narrative describing the action taken to correct the discrepancy’ [15, pp. 4-5-6]. This guidance for free text entries, found in AFTO 00-20-2 – Maintenance Data Documentation, assumes that Servicemembers write comprehensive maintenance entries. *The critical assumption as applied to the study is that Servicemembers write free text narratives for discrepancies and corrective actions that incorporate the object, maintenance action, malfunction, and if the malfunction is corrosion-related with sufficient detail so that the respective maintenance code keyword(s) will be reflected in the narrative.* Army publications (DA PAM 750-8 – The Army Maintenance Management System (TAMMS) User Manual), Navy publications (COMNAVAIRFORINST 4790.2C – The Naval Aviation Maintenance Program) and Marine Corps publications (MCTP 3-40E – Marine Corps Maintenance Operations) all have similar language and expectations for free text entries [11] [12] [13]. It is fair to assume that all free text entries are as comprehensive as possible, although it certainly varies by individual, by supervisor, and by organization.

The second assumption for this study assumes that the highest Jaro-Winkler similarity score is the best match between the coded data field and the free text data field. It is possible that cases exist where the highest scoring pair should have yielded a non-match, and a lower scoring pair should have been classified as a match. This could occur because of misspellings in the records, shorthand/acronyms used, or longer phrases used to achieve a higher score. This assumption does not impact a significant number of records since a combinatorial pairwise comparison is done between tokens, so if two of the same/similar words appear between coded and free text fields, the similarity score will be very high and be classified as a match.

The third assumption is that 70% of the ‘potential match’ records are classified as ‘match.’ This assumption took any records with a Jaro-Winkler similarity score from [0.79, 0.92) and multiplied the total frequency by 0.70 to find the match proportion to classify as a ‘match.’ This was based on the analyst reviewing 100 ‘potential match’ records and finding that about 70% of ‘potential match’ records deserved to be considered as a ‘match.’ Assessing all ‘potential match’ records would have been cumbersome; therefore, this assumption was necessary to calculate the overall estimated consistency match score.

Section 4 Findings

4.1. Findings: Navy and Air Force H-60 Maintenance Data Quality

Using the Jaro-Winkler Algorithm to compare, classify, and evaluate Navy and Air Force H-60 coded maintenance data fields to free text data fields yielded information and insights that will be explained in Section 4.1. This section will cover the findings for the Navy H-60 maintenance records (organizational and depot), then cover the findings for the Air Force H-60 maintenance records (organizational and depot). The findings answer the hypotheses posed in Section 2.8.1 and 2.8.2 as well as address maintenance record completeness for each data source.

For Navy organizational H-60 maintenance records, 147,795 records were assessed for consistency and completeness for 'object receiving maintenance.' When comparing the WUC field to the free text field, an estimated 60.98% of records had Jaro-Winkler similarity scores that indicated 'match.' 52.53% of records had a Jaro-Winkler similarity score of [0.92 to 1.0] (match) and 12.07% of records had a Jaro-Winkler similarity score of [0.79 to 0.92] (potential match). Applying the 0.70 match proportion to the 'potential match' classification yields the 60.98% estimated object consistency match score. The 95% confidence interval for the comparison yields a range of 59.03% to 62.93%. Figure 9 highlights the distribution of 'match,' 'potential match,' and 'non-match' records. Of specific note, many records (17,667) had a Jaro-Winkler similarity score of 0.0. This relates to the completeness of the dataset. While an insignificant number of WUC or free text fields were completely blank (4 records of 147,795 total), 11.95% of WUCs were not found within the COMNAVAIRFORINST 4790.2C & DECKPLATE reference sources. This indicates that users/maintainers are inputting WUCs that are not in the source documents. Another explanation is that updates to the references have occurred but are not widely disseminated (e.g, DECKPLATE does not reflect the updated WUCs).

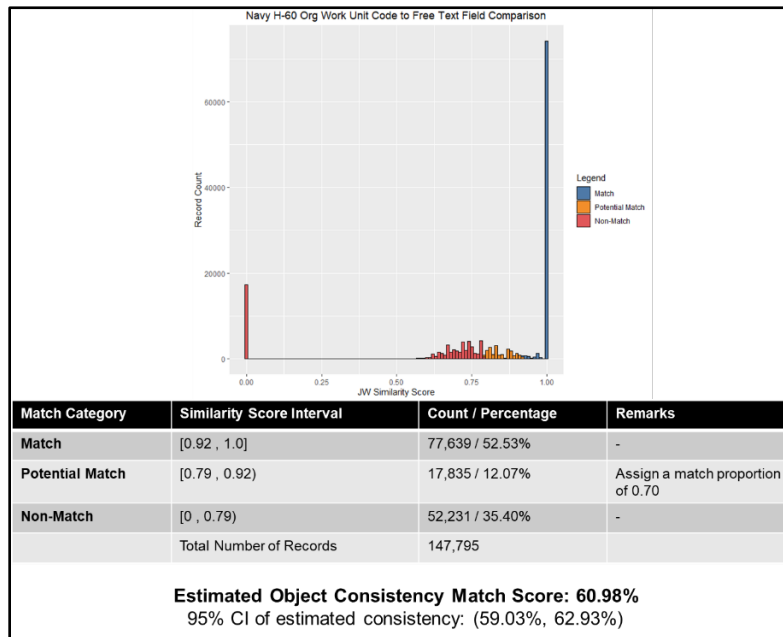


Figure 9: Navy H-60 Organizational Maintenance Data Part Code to Free Text Comparison

For Navy organizational H-60 maintenance records, 147,795 records were assessed for consistency and completeness for 'maintenance action taken.' When comparing the Action Taken Code field to the free text field, an estimated 84.05% of records had Jaro-Winkler similarity scores that indicated 'match.' 76.13% of records had a Jaro-Winkler similarity score of [0.92 to 1.0] (match) and 11.31% of records had a Jaro-Winkler similarity score of [0.79 to 0.92) (potential match). Applying the 0.70 match proportion to the 'potential match' classification yields the 84.05% estimated action taken consistency match score. The 95% confidence interval for the comparison yields a range of 82.99% to 85.11%. Figure 10 highlights the distribution of 'match,' 'potential match,' and 'non-match' records. The dataset was very complete, with all AT Codes matching the reference documents and 7 blank coded fields out of 147,795. Of all the H-60 maintenance records assessed, the Navy organizational H-60 data had the highest consistency and had an extremely high level of completeness.

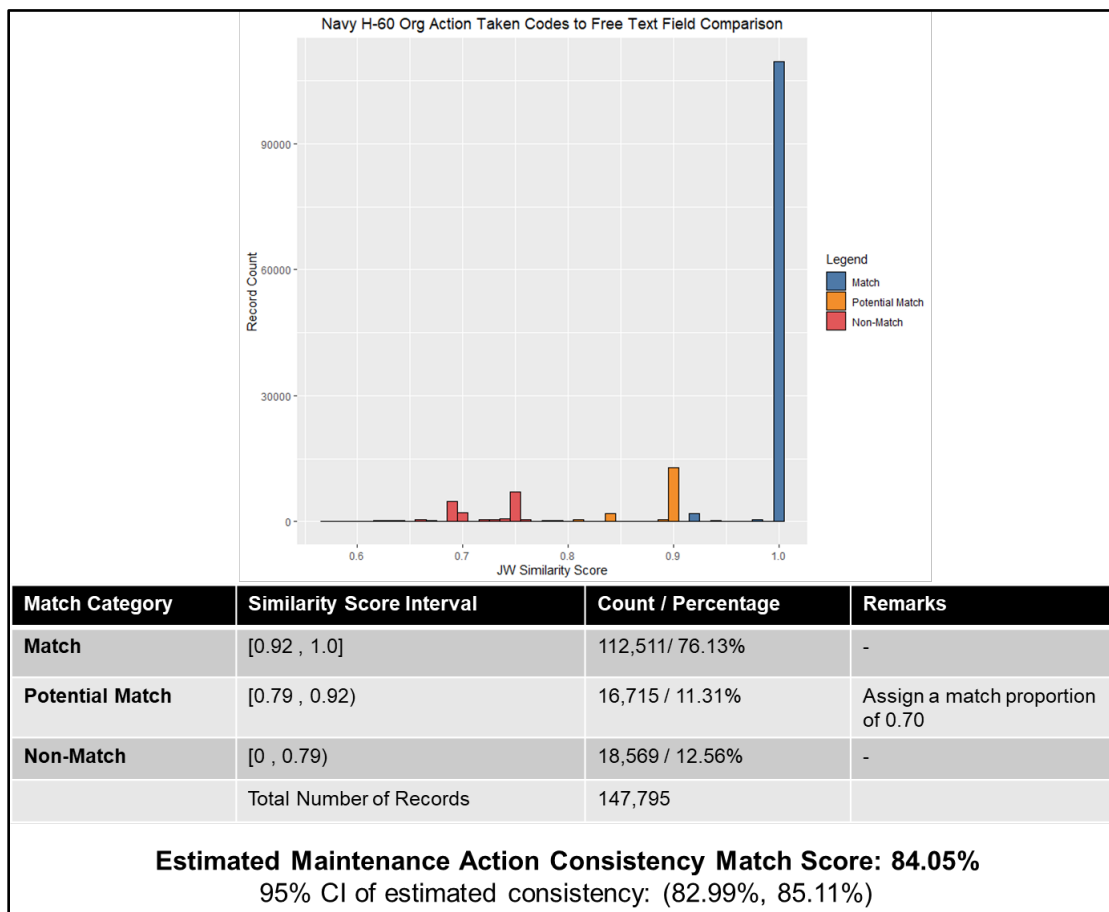


Figure 10: Navy H-60 Organizational Maintenance Data Action Taken Code to Free Text Comparison

For Navy organizational H-60 maintenance records, 147,795 records were assessed for consistency and completeness for 'maintenance malfunction.' When comparing the MALCODE field to the free text field, an estimated 78.52% of records had Jaro-Winkler similarity scores that indicated 'match.' 59.29% of records had a Jaro-Winkler similarity score of [0.92 to 1.0] (match) and 27.47% of records had a Jaro-Winkler

similarity score of [0.79 to 0.92) (potential match). Applying the 0.70 match proportion to the 'potential match' classification yields the 78.52% estimated malfunction consistency match score. The 95% confidence interval for the comparison yields a range of 76.68% to 80.36%. Figure 11 highlights the distribution of 'match,' 'potential match,' and 'non-match' records. The Navy organizational H-60 maintenance data only have 2 records of a possible 147,795 records with MALCODES not found within the COMNAVAIRFORINST 4790.2C & DECKPLATE reference sources. This indicates a very complete maintenance record set. The consistency for the Navy H-60 organizational data was the highest of all H-60 data assessed (78.52% match score). This indicates that the malfunction/discrepancy that the users/maintainers are describing in the free text is the same malfunction/discrepancy that they entered a MALCODE for, which is a promising finding.

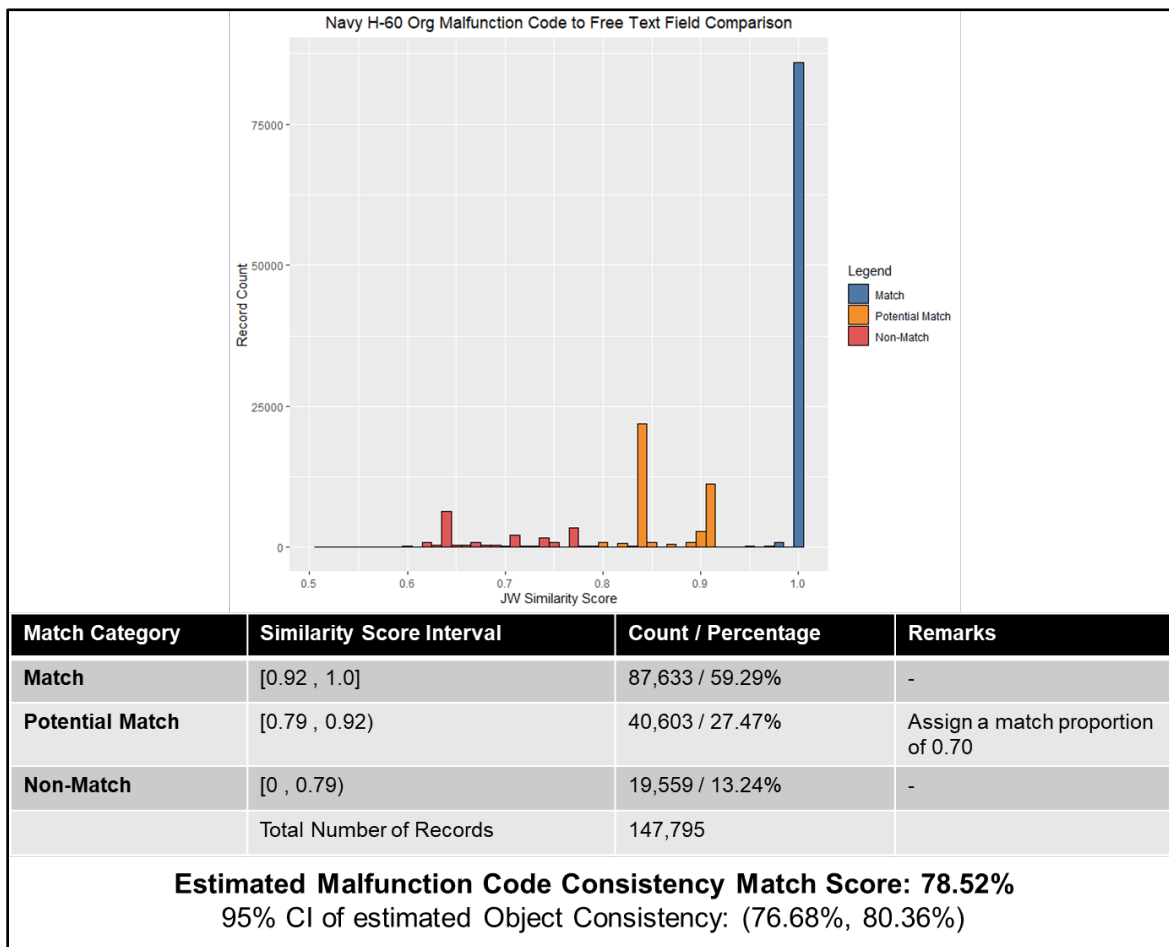


Figure 11: Navy H-60 Organizational Maintenance Data Malfunction Code to Free Text Comparison

For Navy organizational H-60 maintenance records, 147,795 records were assessed based on MALCODE to determine if a record was corrosion-related. 128,727 of the records had a 'corrosion-positive' MALCODE, and 19,068 of the records had a 'corrosion-negative' MALCODE. Figure 12 depicts the distribution of

'corrosion-positive' records that had free text narratives that included corrosion keywords from the MADW Corrosion Dictionary. 82.23% of records with corrosion-positive MALCODEs had corrosion-related free text. Similarly, Figure 13 depicts the distribution of 'corrosion-negative' records that had free text narratives that included corrosion keywords from the MADW Corrosion Dictionary. 7.21% of records with corrosion-negative MALCODEs still had corrosion-related free text. Figure 14 then combines these results to develop the error matrix and performance metrics for Navy organizational H-60 maintenance data. The Navy organizational maintenance records are fairly accurate, with an accuracy score of 83.60%. The records are also fairly precise (82.24%), meaning that 82.24% of the time, a record with a 'corrosion-positive' MALCODE will also have free text with corrosion-related keywords (based on coded and free text field consistency). Recall is extremely high (98.72%), indicating that of records with corrosion-positive free text, 98.72% of those records are properly classified. Finally, the F_1 score (weighted average of precision and recall that penalizes extreme values) is 89.73%. This is a very high F_1 score for this type of natural language processing classification analysis. It is fair to say that for Navy organizational H-60 maintenance data, records are very consistent based on several performance metrics. There is an issue with completeness, based on 11.95% of WUCs not being found in the reference documents.

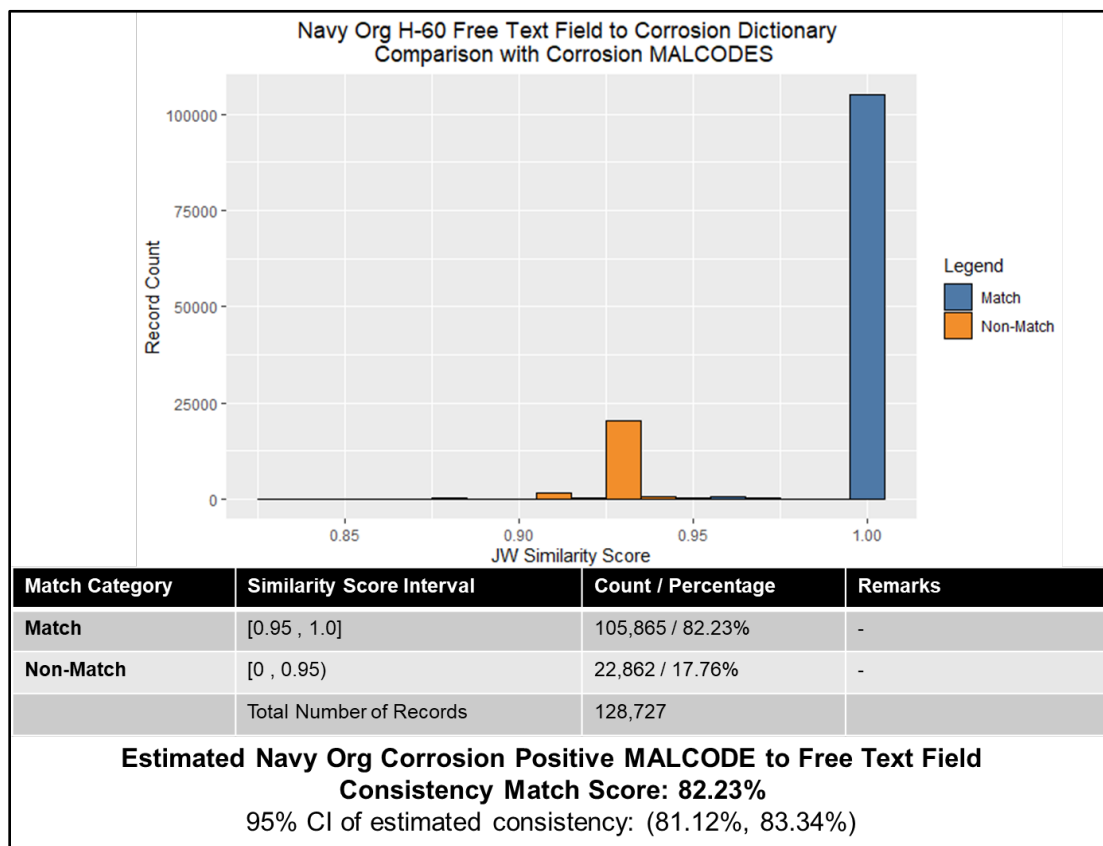


**Figure 12: Navy H-60 Organizational Corrosion-Positive Malfunction Code
to Free Text Comparison**

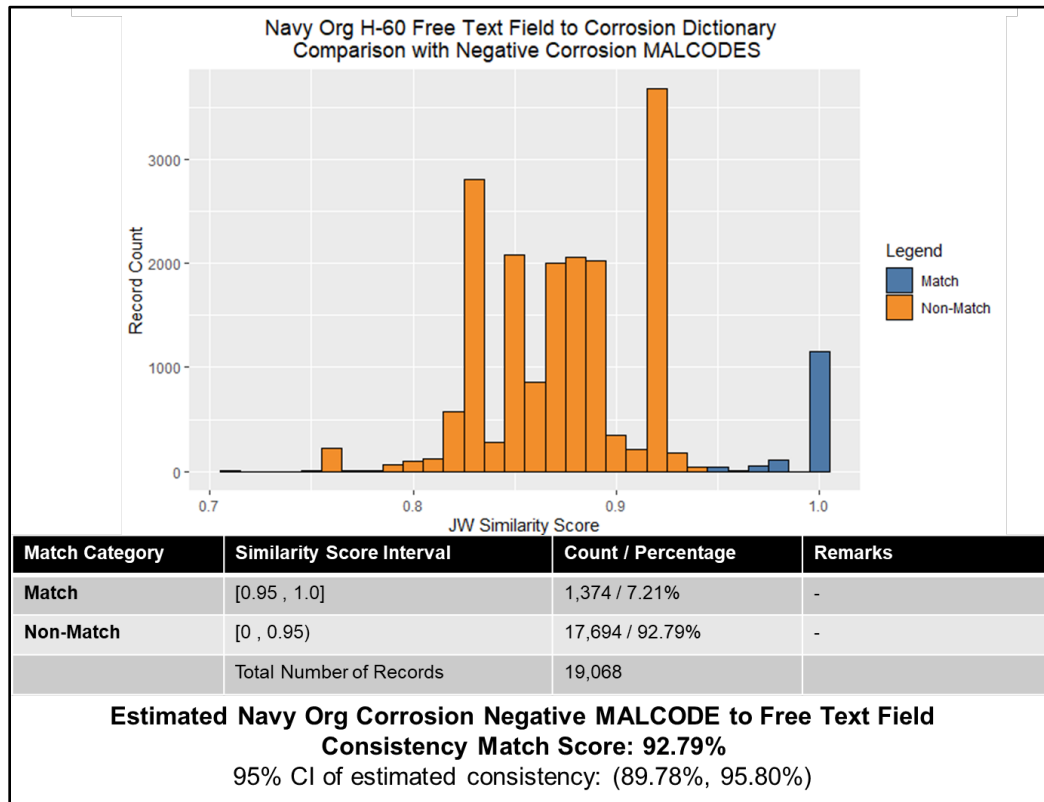


Figure 13: Navy H-60 Organizational Corrosion-Negative Malfunction Code to Free Text Comparison

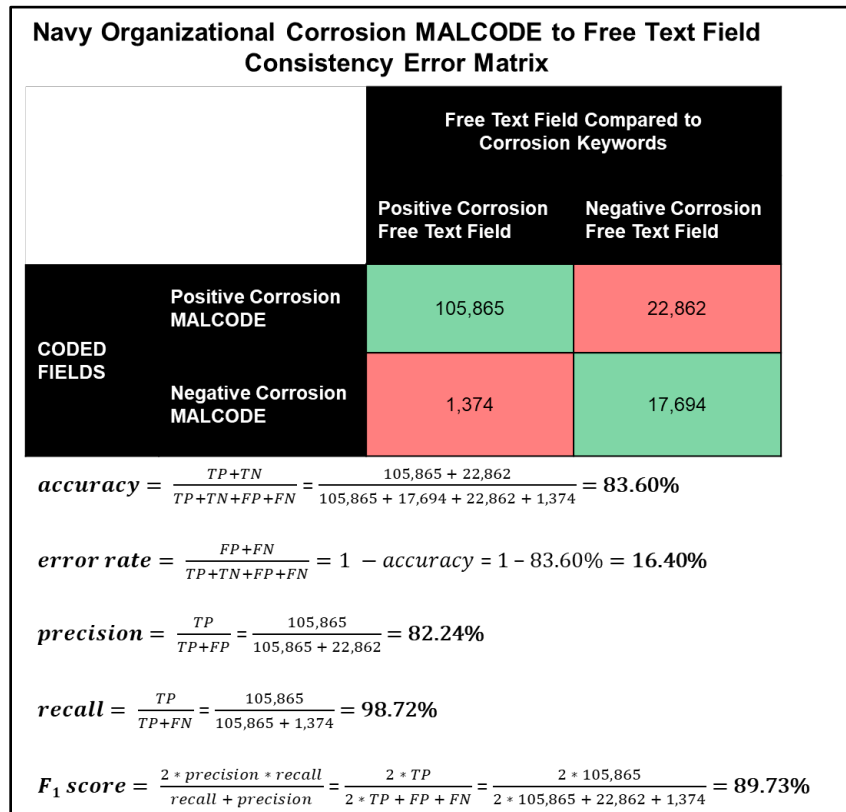


Figure 14: Navy H-60 Organizational Corrosion-Consistency Error Matrix and Performance Metrics

For the Navy depot H-60 maintenance records, Air Force organizational H-60 maintenance records, and Air Force depot H-60 maintenance records, the summaries will be included in this section. For all corresponding graphics, refer to Appendix F.

For Navy depot H-60 maintenance records, 19,838 records were assessed for consistency and completeness for 'object receiving maintenance.' When comparing the WUC field to the free text field, an estimated 49.49% of records had Jaro-Winkler similarity scores that indicated 'match.' 34.63% of records had a Jaro-Winkler similarity score of [0.92 to 1.0] (match) and 21.23% of records had a Jaro-Winkler similarity score of [0.79 to 0.92] (potential match). Applying the 0.70 match proportion to the 'potential match' classification yields the 49.49% estimated object consistency match score. The 95% confidence interval for the comparison yields a range of 41.09% to 57.89%. Figure 21 highlights the distribution of 'match,' 'potential match,' and 'non-match' records for the object receiving maintenance. The 49.49% object consistency match score is lower than the 60.98% match score for Navy organizational H-60 records. When assessing completeness, 816 records were blank, and 158 records had WUCs not found in the source references (COMNAVAIRFORINST 4790.2C & DECKPLATE). This equates to 4.91% of WUCs being unusable for comparison testing for the record set. This reflects an improvement compared to the Navy organizational H-60 data.

For Navy depot H-60 maintenance records, there were no data fields to assess action taken for consistency or completeness. This is a function of the way maintenance data is collected at the three major depot



maintenance locations (Fleet Readiness Center South East in Jacksonville, FL, Fleet Readiness Center Mid-Atlantic in Norfolk, VA, and Fleet Readiness Center South West in North Island, CA). Without relevant data fields, action taken consistency and completeness was not assessed.

For Navy depot H-60 maintenance records, 19,838 records were assessed for consistency and completeness for 'maintenance malfunction.' When comparing the MALCODE field to the free text field, an estimated 67.66% of records had Jaro-Winkler similarity scores that indicated 'match.' 58.69% of records had a Jaro-Winkler similarity score of [0.92 to 1.0] (match) and 12.81% of records had a Jaro-Winkler similarity score of [0.79 to 0.92] (potential match). Applying the 0.70 match proportion to the 'potential match' classification yields the 67.66% estimated malfunction consistency match score. The 95% confidence interval for the comparison yields a range of 63.59% to 71.73%. Figure 22 highlights the distribution of 'match,' 'potential match,' and 'non-match' records. The Navy depot H-60 maintenance data only have 121 records of a possible 19,838 records with MALCODES not found within the COMNAVAIRFORINST 4790.2C & DECKPLATE reference sources and an additional 500 blank MALCODES. This indicates a fairly complete maintenance record set (3.13% of records with completeness issues). Overall, the consistency and completeness for Navy depot H-60 maintenance records for malfunction does not exhibit any major issues.

For Navy depot H-60 maintenance records, 18,864 records were assessed based on MALCODE to determine if a record was corrosion-related. All 18,864 of the records had a 'corrosion-positive' MALCODE, which was unusual. This is a function of the data provided, and further analysis is required with a more comprehensive dataset to thoroughly evaluate data quality for Navy depot H-60 maintenance. Figure 29 depicts the distribution of 'corrosion-positive' records that had free text narratives that included corrosion keywords from the MADW Corrosion Dictionary. 92.66% of records with corrosion-positive MALCODEs had corrosion-related free text. Figure 30 then combines these results to develop the error matrix and performance metrics for Navy depot H-60 maintenance data. The Navy organizational maintenance records are very accurate, with an accuracy score of 92.66%. The records are also very precise (92.66%), meaning that 92.66% of the time, a record with a 'corrosion-positive' MALCODE will also have free text with corrosion-related keywords (based on coded and free text field consistency). Recall and F₁ score are both extremely high (100.0% and 96.19% respectively) but should not be considered because of the lack of records with 'corrosion-negative' MALCODES. Without this category of record present, recall and F₁ score should be dismissed and not used to compare across data sources. It is fair to say that for Navy depot H-60 maintenance data, if a record has a corrosion-related MALCODE, then the record is very likely (92.66% of the time) to be corrosion-related. Any other conclusions related to performance metrics are difficult to assess based on the type of data available.

For Air Force organizational H-60 maintenance records, 50,760 records were assessed for consistency and completeness for 'object receiving maintenance.' When comparing the WUC field to the free text field, an estimated 74.49% of records had Jaro-Winkler similarity scores that indicated 'match.' 67.40% of records had a Jaro-Winkler similarity score of [0.92 to 1.0] (match) and 10.13% of records had a Jaro-Winkler similarity score of [0.79 to 0.92] (potential match). Applying the 0.70 match proportion to the 'potential match' classification yields the 74.49% estimated object consistency match score. The 95% confidence interval for the comparison yields a range of 72.19% to 76.79%. Figure 23 highlights the distribution of 'match,' 'potential match,' and 'non-match' records. When assessing completeness, all maintenance records had a WUC and free text field completed; however, 859 records (1.69%) of WUCs were not found in the AFTO 1H-60(H)G-06: Organizational Maintenance Manual. This indicates that a small number of entries by users/maintainers include WUCs that are not in the source documents. Overall, the consistency and completeness of the Air Force organizational H-60 maintenance records for object receiving maintenance are satisfactory.



For Air Force organizational H-60 maintenance records, 50,760 records were assessed for consistency and completeness for 'maintenance action taken.' When comparing the Action Taken Code field to the free text field, an estimated 59.09% of records had Jaro-Winkler similarity scores that indicated 'match.' 45.67% of records had a Jaro-Winkler similarity score of [0.92 to 1.0] (match) and 19.17% of records had a Jaro-Winkler similarity score of [0.79 to 0.92) (potential match). Applying the 0.70 match proportion to the 'potential match' classification yields the 59.09% estimated action taken consistency match score. The 95% confidence interval for the comparison yields a range of 82.99% to 85.11%. Figure 24 highlights the distribution of 'match,' 'potential match,' and 'non-match' records. The dataset was very complete, with zero blank AT Code fields or free text fields. There were 192 records (0.38%) of records that had AT codes not found in the reference documents. Of all the H-60 maintenance records assessed, the Air Force organizational H-60 data had the lowest consistency score but did have a high level of completeness.

For Air Force organizational H-60 maintenance records, 50,760 records were assessed for consistency and completeness for 'maintenance malfunction.' When comparing the MALCODE field to the free text field, an estimated 64.90% of records had Jaro-Winkler similarity scores that indicated 'match.' 59.14% of records had a Jaro-Winkler similarity score of [0.92 to 1.0] (match) and 8.23% of records had a Jaro-Winkler similarity score of [0.79 to 0.92) (potential match). Applying the 0.70 match proportion to the 'potential match' classification yields the 64.90% estimated malfunction consistency match score. The 95% confidence interval for the comparison yields a range of 62.48% to 67.32%. Figure 25 highlights the distribution of 'match,' 'potential match,' and 'non-match' records. The Air Force organizational H-60 maintenance data have 793 records of a possible 50,760 records (1.56%) with MALCODES not found within the AFTO 1H-60(H)G-06: Organizational Maintenance Manual reference source. This indicates a complete maintenance record set. The malfunction consistency score for Air Force organizational H-60 maintenance data was the lowest of the four record sets, indicating that either the narrative descriptions are too short/incomplete, the MALCODES do not represent the actual discrepancies, or the MALCODE selected is not consistent with the free text narrative.

For Air Force organizational H-60 maintenance records, 50,760 records were assessed based on MALCODE to determine if a record was corrosion-related. 2,304 of the records had a 'corrosion-positive' MALCODE, and 48,456 of the records had a 'corrosion-negative' MALCODE. Figure 31 depicts the distribution of 'corrosion-positive' records that had free text narratives that included corrosion keywords from the MADW Corrosion Dictionary. 87.46% of records with corrosion-positive MALCODEs had corrosion-related free text. Similarly, Figure 32 depicts the distribution of 'corrosion-negative' records that had free text narratives that included corrosion keywords from the MADW Corrosion Dictionary. 3.00% of records with corrosion-negative MALCODEs still had corrosion-related free text. Figure 33 then combines these results to develop the error matrix and performance metrics for Navy organizational H-60 maintenance data. The Air Force organizational maintenance records are extremely accurate, with an accuracy score of 96.57%. The records are also precise (87.46%), meaning that 87.46% of the time, a record with a 'corrosion-positive' MALCODE will also have free text with corrosion-related keywords (based on coded and free text field consistency). Recall is fairly low (58.12%), indicating that of records with corrosion-related free text, 58.12% of those records are properly classified; however, this low score is partly a function of having such a large percentage of all records (95.56%) of all records not having corrosion-positive MALCODES. Finally, the F₁ score (weighted average of precision and recall that penalizes extreme values) is 69.83%. This is a fairly low F₁ score but has the same issue as recall because of the limited sample of corrosion-positive MALCODE records relative to the entire sample. Overall, the Air Force organizational H-60 records are very accurate, very consistent, and very complete.

For Air Force depot H-60 maintenance records, 13,199 records were assessed for consistency and completeness for 'object receiving maintenance.' Not all records from the multiple Air Force maintenance



depots collected/provided WUCs that could be used for comparison, which is why the entire record set is not included. When comparing the WUC field to the free text field, an estimated 96.19% of records had Jaro-Winkler similarity scores that indicated 'match.' 92.84% of records had a Jaro-Winkler similarity score of [0.92 to 1.0] (match) and 4.79% of records had a Jaro-Winkler similarity score of [0.79 to 0.92] (potential match). Applying the 0.70 match proportion to the 'potential match' classification yields the 96.19% estimated object consistency match score. The 95% confidence interval for the comparison yields a range of 93.67% to 98.71%. Figure 26 highlights the distribution of 'match,' 'potential match,' and 'non-match' records. When assessing completeness, all maintenance records had a WUC and free text field completed. Unlike other record sets, all WUCs corresponded to the reference source, so there were zero completeness issues across the entire maintenance record set for object receiving maintenance. Overall, the consistency and completeness of the Air Force depot H-60 maintenance records for object receiving maintenance are outstanding.

For Air Force depot H-60 maintenance records, 29,484 records were assessed for consistency and completeness for 'maintenance action taken.' Not all records from the multiple Air Force maintenance depots collected/provided Action Taken Codes that could be used for comparison, which is why the entire record set is not included. When comparing the Action Taken Code field to the free text field, an estimated 79.35% of records had Jaro-Winkler similarity scores that indicated 'match.' 67.39% of records had a Jaro-Winkler similarity score of [0.92 to 1.0] (match) and 17.08% of records had a Jaro-Winkler similarity score of [0.79 to 0.92] (potential match). Applying the 0.70 match proportion to the 'potential match' classification yields the 79.35% estimated action taken consistency match score. The 95% confidence interval for the comparison yields a range of 76.22% to 82.48%. Figure 27 highlights the distribution of 'match,' 'potential match,' and 'non-match' records. When assessing completeness, all maintenance records had a AT Code and free text field completed. Unlike other record sets, all AT Codes corresponded to the reference source, so there were zero completeness issues across the entire maintenance record set for maintenance action taken.

For Air Force depot H-60 maintenance records, 17,216 records were assessed for consistency and completeness for 'maintenance malfunction.' When comparing the MALCODE field to the free text field, an estimated 71.91% of records had Jaro-Winkler similarity scores that indicated 'match.' 44.59% of records had a Jaro-Winkler similarity score of [0.92 to 1.0] (match) and 39.03% of records had a Jaro-Winkler similarity score of [0.79 to 0.92] (potential match). Applying the 0.70 match proportion to the 'potential match' classification yields the 71.91% estimated malfunction consistency match score. The 95% confidence interval for the comparison yields a range of 70.11% to 73.71%. Figure 28 highlights the distribution of 'match,' 'potential match,' and 'non-match' records. The Air Force depot H-60 maintenance data have 231 records of a possible 17,216 records (1.56%) with MALCODES not found within the AFTO 1H-60(H)G-06: Organizational Maintenance Manual reference source. This indicates a complete maintenance record set. The malfunction consistency score for Air Force organizational H-60 maintenance data was the second highest of the four record sets assessed, indicating that the narrative descriptions generally match the MALCODES.

For Air Force depot H-60 maintenance records, 17,216 records were assessed based on MALCODE to determine if a record was corrosion-related. 8,543 of the records had a 'corrosion-positive' MALCODE, and 8,673 of the records had a 'corrosion-negative' MALCODE. Figure 34 depicts the distribution of 'corrosion-positive' records that had free text narratives that included corrosion keywords from the MADW Corrosion Dictionary. 86.14% of records with corrosion-positive MALCODEs had corrosion-related free text. Similarly, Figure 35 depicts the distribution of 'corrosion-negative' records that had free text narratives that included corrosion keywords from the MADW Corrosion Dictionary. 9.95% of records with corrosion-negative MALCODEs still had corrosion-related free text. Figure 36 then combines these results to develop the error



matrix and performance metrics for Navy organizational H-60 maintenance data. The Air Force organizational maintenance records are very accurate, with an accuracy score of 88.11%. The records are also precise (86.14%), meaning that 86.14% of the time, a record with a 'corrosion-positive' MALCODE will also have free text with corrosion-related keywords (based on coded and free text field consistency). Recall is also high (89.50%), indicating that of records with corrosion-related free text, 89.50% of those records are properly classified. Unlike other datasets, the distribution between corrosion-positive and corrosion-negative are evenly distributed. This allows use to have more confidence in the recall and precision metrics. Finally, the F_1 score (weighted average of precision and recall that penalizes extreme values) is 87.79%. This is a high F_1 score and reflects that the Air Force depot H-60 maintenance data demonstrates both precision and good recall. Overall, the Air Force depot H-60 records are very accurate, very consistent, and very complete.

Having assessed the Navy and Air Force H-60 maintenance records (both organizational and depot-level), a few key findings emerge related to the quality of the data. No single record type is dominant when considering the multiple hypotheses that were assessed. For the object receiving maintenance, the Air Force H-60 depot-level maintenance records scored the highest, with a 96.19% object match consistency score. For maintenance action taken, Navy H-60 organizational maintenance records scored the highest, with an action taken match consistency score of 84.05%. For maintenance discrepancy, Navy H-60 organizational maintenance records again scored the highest, with a malfunction match consistency score of 78.52%. When comparing corrosion-coded fields to free text fields, comparing F_1 scores would be ideal to compare across record sets; however, with the significantly different proportions of corrosion-positive MALCODES, accuracy is the most reasonable metric to use for comparison across datasets. Precision and recall have the same issue as F_1 score since the distribution of records are disparate across the Service organizational and depot-level records. Navy and Air Force organizational and depot-level maintenance records exhibit high accuracy levels (all exceed 83%) (see Table 12). Air Force H-60 organizational maintenance records have the highest accuracy when comparing coded fields to free text fields and classifying both the coded fields and free text fields into corrosion-positive and corrosion-negative categories. Overall record completeness is higher for the Air Force maintenance records than for the Navy maintenance records (fewer blanks or coded fields not in TMs). Object data fields had the lowest record completeness for both Services of all categorizes that were evaluated. This is likely because WUCs are manually entered and are five to eight characters in length and there are several thousand possible WUCs (7,399 WUCs for Navy, 2,319 WUCs for Air Force), which introduces more opportunity for input errors from the users. The maintenance action taken fields are very complete since this field only requires two characters and there are relatively few possible fields (29 AT Codes for Navy, 33 AT Codes for Air Force). Although some malfunction field input errors exist, overall, more than 98% of malfunction records are complete with no quality issues (322 MALCODES for Navy, 225 MALCODES for Air Force). An issue that is clear is that the data samples may not be representative. For example, the Navy H-60 depot record set includes only corrosion-related MALCODES. This makes it impossible to assess other maintenance actions and generalize the insights about record quality, since only one type of record classification.

**Table 12: Results for H-60 Maintenance Records**

Record Type	Accuracy for Corrosion Match Scoring	F ₁ for Corrosion Match Scoring
Navy Organizational H-60 Maintenance Data	83.60%	89.73%
Navy Depot H-60 Maintenance Data*	92.66%*	96.19%*
Air Force Organizational H-60 Maintenance Data	96.57%	69.83%
Air Force Depot H-60 Maintenance Data	88.11%	87.79%

* Navy Depot H-60 Maintenance Data included all corrosion-positive data records.

4.2. Findings: Best Practices and Recommendations for Corrosion Maintenance

Having assessed Navy and Air Force H-60 maintenance data and using the Army H-60 analysis from the FY20 CPO Study by MAJ Case, comparisons, best practices, and recommendations are identified. This section identifies major differences across the Services in maintenance data collection, then highlights best practices and make recommendations for improving corrosion-related maintenance.

Navy H-60 maintenance data is managed and organized well, with some possible improvements possible for depot maintenance records. Both organizational and depot level records collect all key administrative data (job control number, unit, parent organization, location, etc.). Organizational maintenance records collect all critical maintenance data in two forms, coded and free text. Maintenance object, maintenance action, and maintenance malfunction are all captured via coded data fields. The user/maintainer then also enters a free text narrative in the 'System Reason Description,' 'Discrepancy,' and 'Corrective Action' fields. This replication within the maintenance record was critical for this study to assess record consistency. It allowed for a way to evaluate how "truthful" a record is in relation to the topic area in question. Depot maintenance records obtained from PMA-299 had similar properties but lacked some specificity. Coded fields existed for maintenance object and maintenance malfunction. Free text existed for the 'Discrepancy Description' field. Depot maintenance records lack coded fields that target maintenance actions and lack free text fields that address maintenance or corrective actions. Additional descriptive data would increase the fidelity of depot-level maintenance data. The complexity and quantity of WUCs within the Navy (7,399 codes) may contribute to the issues with record completeness (11.95% incomplete/non-matching WUCs for organizational data). This may also be a contributing factor for the lower object consistency match scores for Navy organizational and depot H-60 maintenance data (60.98% and 49.49% respectively). Navy maintenance consistency scores were higher for action codes and malfunction codes; this may be because of fewer codes available for selection. When reviewing the data collection process for the Navy, a major point to highlight is the Aircraft Delivery Deficiency Report (ADDR). The ADDR collects deficiencies (ADCS records) from inadequate periodic maintenance found at the depot level by aircraft examiners. The ADDR is then sent back to the individual squadron to help the unit identify and improve maintenance quality at the organizational level (OOMA records) [12, pp. 3-1]. This serves as a feedback loop to help improve Navy maintenance at the organizational level. ADDRs capture errors in maintenance records (data quality issues) as well as incorrect maintenance decisions/actions taken. Further analysis is required into the ADDR program to assess its effectiveness, but the concept of feedback from a higher level of maintenance to a lower level of maintenance is noteworthy.



When assessing the maintenance data quality holistically, the Air Force H-60 maintenance data was superior to Navy H-60 maintenance data. Both organizational and depot level records scored well in the consistency and completeness assessments. The Air Force organizational H-60 maintenance records were similar to the Navy in that both collected the required administrative data that allows for analysis by parent organization, unit location, date accepted for maintenance, and other pertinent fields. Air Force organizational maintenance records collect critical maintenance data in two forms, coded and free text. Having specific codes for maintenance object, maintenance action, and maintenance malfunction allows for comparison between the 'Maintenance Action Discrepancy' and 'Maintenance Action Corrective Action' fields. Although the depot level H-60 maintenance records were grouped by location and each was organized slightly differently, the critical data fields were collected, enabling consistency and completeness assessments. Speaking with senior maintenance supervisors for the Air Force highlighted a few key differences in data collection for the Air Force. Each Air Force Servicemember is directly responsible for the data quality relating to maintenance actions that are input to IMDS [14, p. 19]. Work Center supervisors conduct quality checks on data input within their respective work centers. Best practice is to check a sample of data entries daily, but sometimes these checks only occur weekly. Air Force maintenance units all have established Data Integrity Teams (DITs) that are responsible for ensuring the unit has complete and accurate data. These DITs identify and quantify problems within their unit and attempt to identify the root cause of the poor data integrity. DITs send error reports to the respective work centers to help educate maintainers and operators that have repeated data quality issues. Units may also implement training programs to correct data integrity trends [14]. The work of DITs is starting to be automated, allowing for a combination of visual inspection and artificial intelligence tools to classify records and identify errors [35]. DIT reports are sent to the specific squadron of origin. By the end of FY22, Conditions-Based Maintenance Plus (CBM+) will use sensors and algorithms to enable predictive maintenance. This will give maintenance units the ability to predict part failures before they occur for all Air Force platforms [36]. Issues identified through interviews with maintenance SMEs is that IMDS has a character limit, so maintainers sometimes cannot fit a full description of a discrepancy or corrective action into the free text entry. To resolve this issue, acronyms emerge and evolve to create a shorthand, often unique to each unit or location. For example, repair and replace is often entered into IMDS as 'R2.' When conducting this study, such acronyms are not known nor accounted for. The maintenance master sergeant interviewed said that "any action that is performed repetitively will become an IMDS acronym." This shorthand contributed to some decrease in consistency scores. Implementing machine learning algorithms or increasing the character length allowed in IMDS could allow for more complete or more correctly categorized records.

Army H-60 maintenance data was found to be of lower quality than that of both the Navy and Air Force H-60 maintenance data. The Army data had the required administrative fields, but much of the data is still reliant on free text entry from the DA2408-13-1 (Aircraft Inspection and Maintenance Record) form. The Aircraft Notebook System is being fielded, which will integrate well with GCSS-A. Aircraft Notebook should improve record accuracy and reduce the amount of manual input for maintenance record activities. This new data collection process is needed, as the current corrosion tag accuracy for the Army ASAP-RAM H-60 maintenance data was 41.6%. This accuracy score is dramatically lower than the Navy and Air Force H-60 corrosion accuracy scores, which ranged from 83.60% (Navy Organizational) to 96.57% (Air Force Organizational). The associated precision (58.8%), recall (33.3%, and F_1 score (42.5%) for the Army H-60 data were all lower than Navy or Air Force H-60 maintenance data [10]. The Army has already identified some of their issues with maintenance quality, hence the shift to Aircraft Notebook for data collection.

Having discussed the H-60 maintenance collection processes for the Navy, Air Force, and Army, best practices and recommendations can be identified. Best practices focus on establishing feedback loops within the data collection process, establishing appropriate corrosion-related codes, and having engaged



unit leadership. Recommendations focus on improving point of entry data collection through NLP AI tools (predictive text), improving training and standardization for data collection, requiring specific data fields to be mandatory, incorporating other forms of data (images, additional coded fields, etc.), and verifying/updating all maintenance activities and malfunctions into technical manuals and PMCS procedures.

A best practice identified in the corrosion maintenance assessment process was the inclusion of feedback loops to improve data quality. Both the Navy and the Air Force have control systems in place that identify errors in the maintenance data collected and provide feedback to the appropriate unit.

The Navy ADDR program communicates potential deficiencies from maintenance that should have been executed at the organizational level. FRC Aircraft Examiner and Evaluators (E&E) evaluate an aircraft at the depot level and provide detailed descriptions of all unscheduled organizational level discrepancies, generating an initial ADDR. As the aircraft proceeds through a depot level maintenance (DLM) event, supplemental ADDRs are generated to capture additional deficiencies. The Close Out ADDR contains information about the overall cost and delivery of the aircraft for the DLM event, as well as what unscheduled organizational level maintenance deficiencies were corrected or 'Noted But Not Corrected' (NBNC) because of funding issues [37]. Appendix G (ADDR Process Flow Chart) shows the expanded ADDR process that highlights how unscheduled organizational level repairs can be completed at the depot and how the ADDR is communicated back to the owning squadron.

The Air Force DITs focus on ensuring squadron (organizational level) maintenance data is complete and accurate. Like the Navy ADDR program, the intent is to communicate maintenance problems to the point of origin. An additional step that DITs focus on is root cause analysis for poor data quality. Unlike the ADDR program, DITs are comprised of organizational level maintainers who "peer-review" the maintenance entries of other units. A specific requirement of DITs is that "the use of general WUCs/LCNs when a component level WUC/LCN exists will be counted as an error" [14, pp. 99-100]. While that is promising and the right direction for data quality (increased specificity is better for fidelity), the current data does not show this to be the case. Of 57,745 records assessed, 22,429 Air Force organizational H-60 records (38.8%) used the 22000 (Turboshaft Engine) or 11000 (Airframe) WUC. Further emphasis is needed to use more details WUCs.

The Navy and Air Force have done an admirable job of establishing appropriate corrosion-related malfunction codes. While ground vehicles in the Army and Marine Corps have very limited numbers of corrosion-related malfunction codes (<5 dedicated malfunction codes), the Air Force has 10 corrosion-specific malfunction codes, and the Navy has 49 corrosion-specific malfunction codes. This specificity is advantageous to allow maintainers to isolate corrosion to specific failure modes.

The Air Force H-60 senior maintainer (E-8) who was interviewed emphasized the importance of engaged unit leadership. He said that he required Work Center supervisors to conduct daily quality checks of maintainer record entries. This helps to develop the individual maintainer and resolves data quality issues immediately, rather than hoping a DIT catches all errors. There is not a standard for how frequently Work Center supervisors need to conduct quality checks of their records, so some supervisors may only check weekly or bi-weekly.

Recommendations from assessing Navy and Air Force H-60 maintenance records highlight possible areas for improvement. The focus needs to be on improving data collection at point of entry through AI tools (predictive text). This could include mobile applications to document maintenance actions that limit free text entry and incorporate other forms of data, like images of the equipment receiving maintenance. Descriptions of maintenance work done could be compiled from PMCS procedures and the equipment technical manuals



such that NLP AI systems could complete the input of a user as they begin to describe the maintenance action through voice or text. This technology already exists in the civilian sector and could be adapted for use in the military.

A second key recommendation is to improve user/maintainer training and standardize data collection. Different units and locations have vastly different standards for what is an acceptable maintenance description, how detailed a malfunction code or WUC should be, and what shorthand/acronym is included in a description. When each organization has its own language, deciphering at the Service or OSD level is challenging. By standardizing what is acceptable for a free text description, better insights can be drawn from maintenance records. Another issue that training needs to address is increasing the specificity of maintenance records. WUCs and MALCODES should be as specific as possible, but often generic codes are used for convenience or lack of user knowledge.

A final recommendation for data quality relates to ensuring data is complete and interpretable. Maintenance forms should have some specific (critical) mandatory fields. This would eliminate some of the completeness issues identified within the maintenance records. Additionally, drop-down/autofill features for completing maintenance forms could ensure that coded fields align with current technical manuals and regulations. There is also a possibility to limit the amount of free text entries and instead include more meaningful images or voice-to-text data. Machine Learning image recognition and classification algorithms are rapidly improving and could be used to tag maintenance issues with a specific part code or discrepancy code.

4.3. Findings: Army and Marine Corps HMMWV Maintenance Data Quality

Using the Jaro-Winkler Algorithm to compare, classify, and evaluate HMMWV coded maintenance data fields to free text data fields yielded information and insights that will be explained in Section 4.3. This section will cover the findings for the Army HMMWV maintenance records, then address the findings for the Marine Corps HMMWV maintenance records. Both sets of data comprise organizational maintenance records based on data availability. The findings address the hypotheses posed in Section 2.8.3 and 2.8.4 as well as address maintenance record completeness for each data source.

For Army organizational HMMWV maintenance records, 13,243 records were assessed for consistency and completeness for 'object receiving maintenance.' When comparing the 'part nomenclature' field to the free text field, an estimated 79.06% of records had Jaro-Winkler similarity scores that indicated 'match.' 37.05% of records had a Jaro-Winkler similarity score of [0.92 to 1.0] (match) and 60.01% of records had a Jaro-Winkler similarity score of [0.79 to 0.92] (potential match). Applying the 0.70 match proportion to the 'potential match' classification yields the 79.06% estimated object consistency match score. The 95% confidence interval for the comparison yields a range of 75.05% to 83.07%. Figure 15 highlights the distribution of 'match,' 'potential match,' and 'non-match' records. The HMMWV maintenance records were fully complete; each record had a Part NSN with corresponding Part Nomenclature. This allowed for straightforward comparison to the Fault Description and Correction Narrative free text fields. Both free text fields were also complete, with no blank fields identified. This indicates that users/maintainers are inputting valid Part NSNs and entering text descriptions for each record.

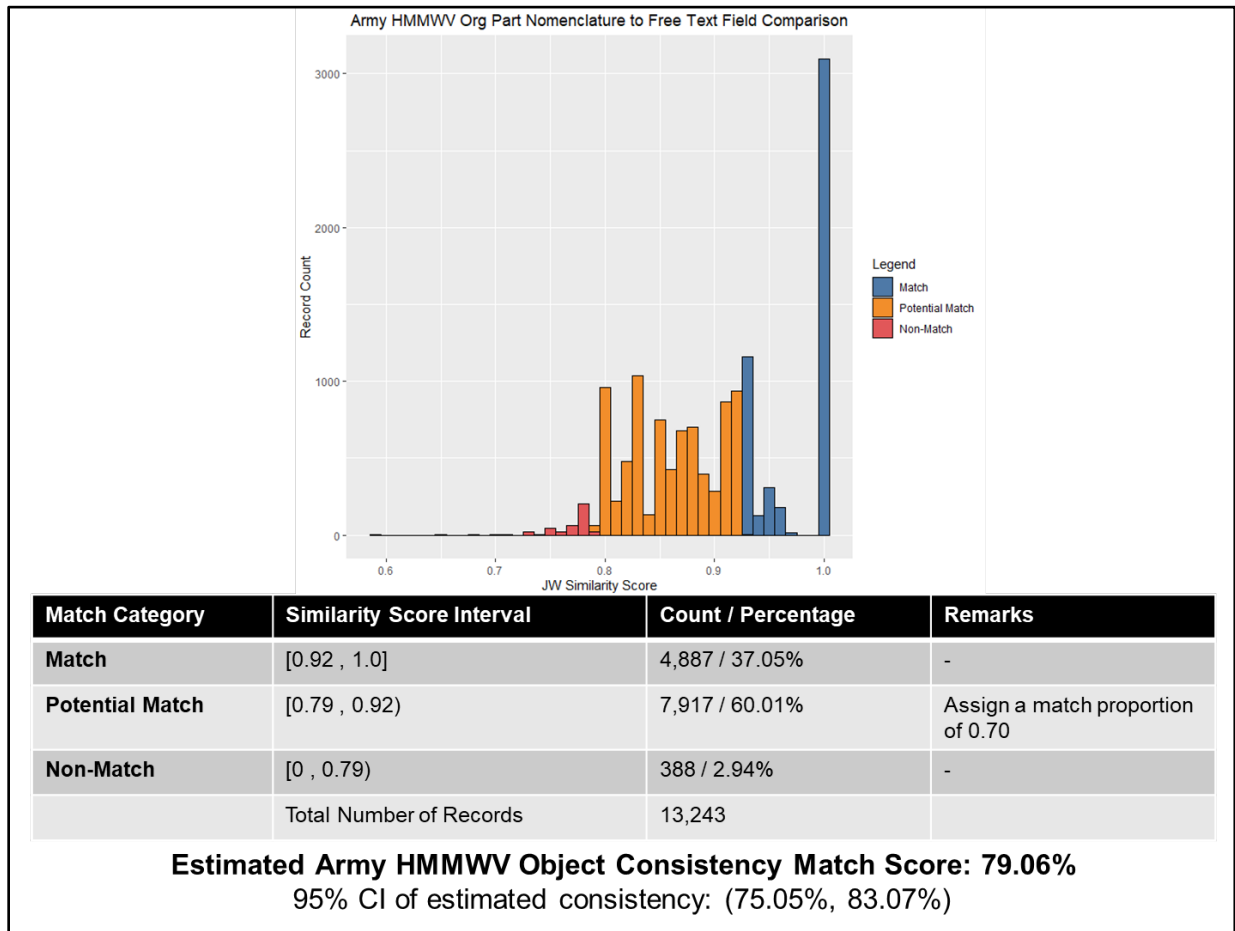


Figure 15: Army HMMWV Organizational Maintenance Data Part Code to Free Text Comparison

For Army organizational HMMWV maintenance records, 13,243 records were assessed for consistency and completeness for 'maintenance action taken.' When comparing the Action Description field to the free text field, an estimated 86.18% of records had Jaro-Winkler similarity scores that indicated 'match.' 86.11% of records had a Jaro-Winkler similarity score of [0.92 to 1.0] (match) and 0.49% of records had a Jaro-Winkler similarity score of [0.79 to 0.92) (potential match). Applying the 0.70 match proportion to the 'potential match' classification yields the 86.11% estimated action taken consistency match score. The 95% confidence interval for the comparison yields a range of 80.37% to 91.99%. Figure 16 highlights the distribution of 'match,' 'potential match,' and 'non-match' records. The HMMWV maintenance records were fully complete; each record had an Action Description selection. These Action Descriptions were succinct words or phrases that a user/maintainer can select (e.g., Adjusted, Replaced, Tested). The Action Description field was compared to the Correction Narrative free text field. Both the coded field and free text field were fully complete, with no blank records found across the dataset. Overall, the maintenance action taken has a high consistency and outstanding completeness for Army HMMWVs.

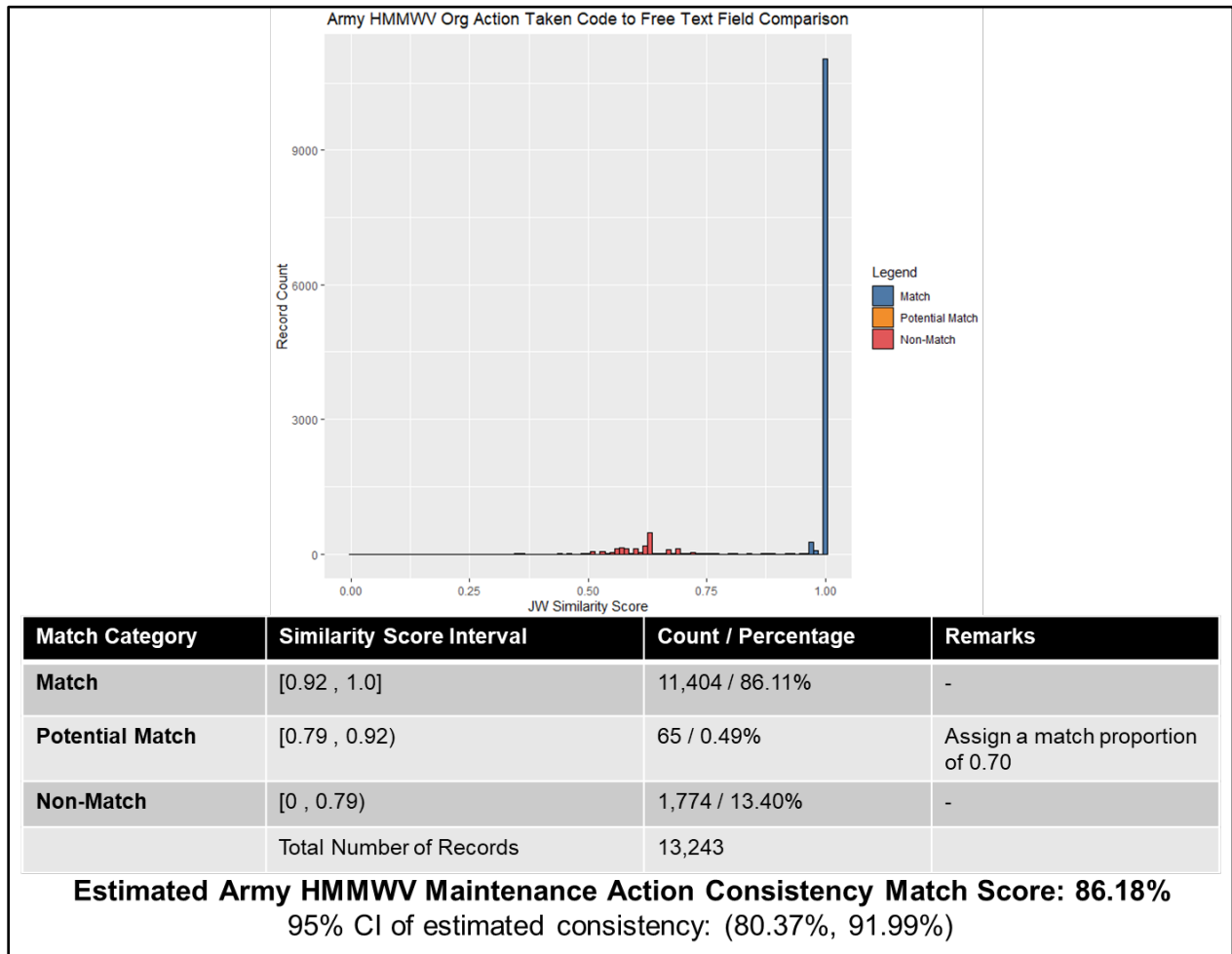


Figure 16: Army HMMWV Organizational Maintenance Data Action Taken Code to Free Text Comparison

For Army organizational HMMWV maintenance records, 13,243 records were assessed for consistency and completeness for 'maintenance malfunction.' When comparing the Failure Code field to the free text fields (Failure Description, Corrective Action, Remarks), an estimated 45.48% of records had Jaro-Winkler similarity scores that indicated 'match.' 34.91% of records had a Jaro-Winkler similarity score of [0.92 to 1.0] (match) and 33.98% of records had a Jaro-Winkler similarity score of [0.79 to 0.92) (potential match). Applying the 0.70 match proportion to the 'potential match' classification yields the 45.48% estimated malfunction consistency match score. The 95% confidence interval for the comparison yields a range of 38.74% to 52.22%. Figure 17 highlights the distribution of 'match,' 'potential match,' and 'non-match' records. The Army organizational HMMWV maintenance data have 125 records of a possible 13,243 records with blank Failure Codes (0.94%). This represents a very complete maintenance record set. The very low match score (34.91%) coupled with a similar potential match score (33.98%) indicates that imprecise language may be used that does not match the Failure Description. Failure Description is mapped directly from the Failure Code. The biggest insight from examining the data is that 9,286 records (70.12%) have a Failure Code of '099' (OTHER/UNKNOWN/BROKEN). This is by far the most used Failure Code,

with the second one being '750' (FAULT DUE TO MISSING COMPONENT(S)) with 1,979 records (14.9%). This indicates that either additional Failure Codes are required that provide more specificity, or additional training is needed so users do not default to Failure Code '099.' To remedy this issue, users should be required to manually enter a failure description to entice them to use the appropriate code or to allow GCSS-A to identify other common malfunctions.

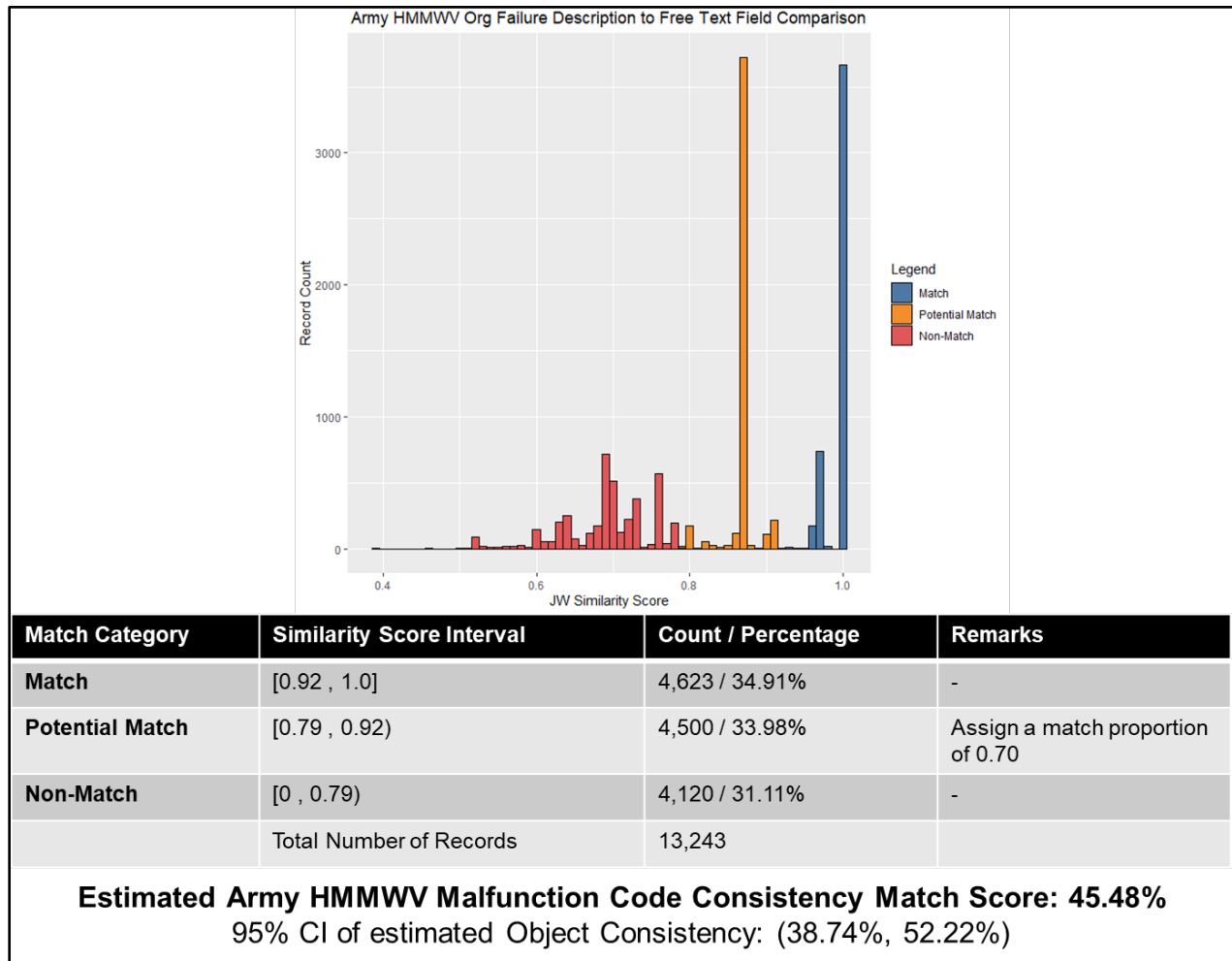


Figure 17: Army HMMWV Organizational Maintenance Data Malfunction Code to Free Text Comparison

For Army organizational HMMWV maintenance records, 13,243 records were assessed based on Failure Code to determine if a record was corrosion-related. Only 51 of the records had a 'corrosion-positive' Failure Code (Code = 170), and 13,192 of the records had a 'corrosion-negative' MALCODE. Figure 18 depicts the distribution of 'corrosion-positive' records that had free text narratives that included corrosion keywords from the MADW Corrosion Dictionary. 96.08% of records with corrosion-positive MALCODEs had corrosion-related free text. Similarly, Figure 19 depicts the distribution of 'corrosion-negative' records that

had free text narratives that included corrosion keywords from the MADW Corrosion Dictionary. 25.25% of records with corrosion-negative Failure Codes still had corrosion-related free text. Figure 20 then combines these results to develop the error matrix and performance metrics for Army organizational HMMWV maintenance data. The Army organizational maintenance records are fairly accurate, with an accuracy score of 74.83%. The records are also very precise (96.08%), meaning that 96.08% of the time, a record with a 'corrosion-positive' Failure Code will also have free text with corrosion-related keywords (based on coded and free text field consistency). Recall is extremely low (1.45%), indicating that of records with corrosion-related free text, 1.45% of those records are properly classified. Finally, the F₁ score (weighted average of precision and recall that penalizes extreme values) is 2.86%, which aligns with the very low recall score. Based on this analysis, there is a high probability that many maintenance records are being mislabeled with their Failure Code. Considering the preponderance of '099' Failure Codes (OTHER/UNKNOWN/BROKEN), it is likely that many of these records should be classified as '170' (CORRODED (METAL) INCLUDES RUSTING). The Failure Codes available for Army ground vehicle maintenance appears to lack specificity which makes it challenging to identify true corrosion-related maintenance records.

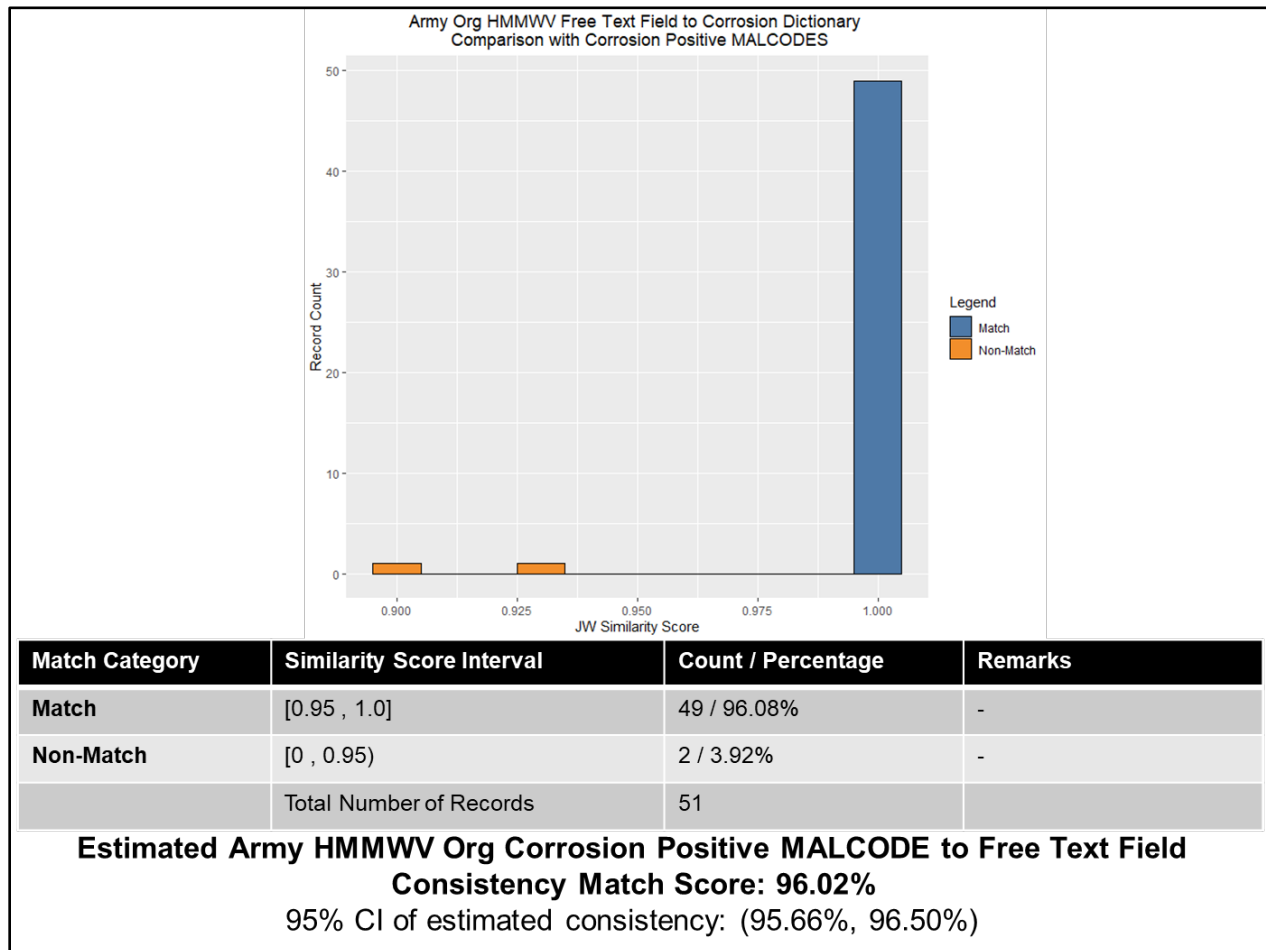
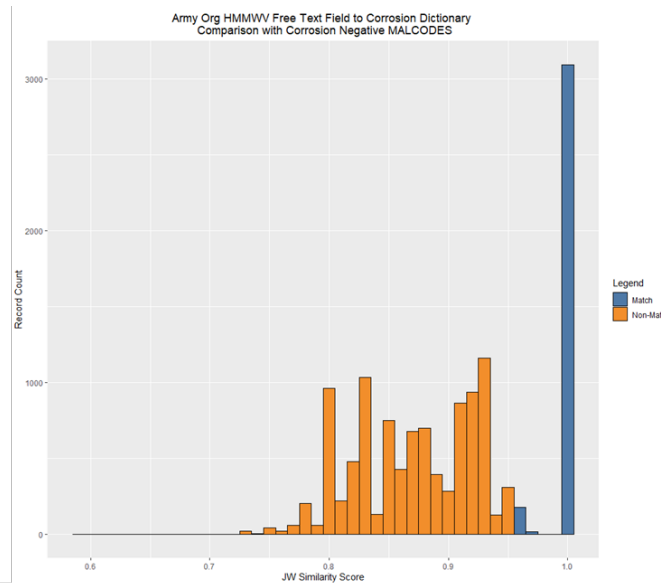


**Figure 18: Army HMMWV Organizational Corrosion-Positive Malfunction Code
to Free Text Comparison**



Match Category	Similarity Score Interval	Count / Percentage	Remarks
Match	[0.95 , 1.0]	3,331/ 25.25%	-
Non-Match	[0 , 0.95)	9,861 / 74.75%	-
Total Number of Records		13,192	

**Estimated Army HMMWV Org Corrosion Negative MALCODE to Free Text Field
Consistency Match Score: 25.25%**
95% CI of estimated consistency: (24.19%, 26.31%)

**Figure 19: Army HMMWV Organizational Corrosion-Negative Malfunction Code
to Free Text Comparison**

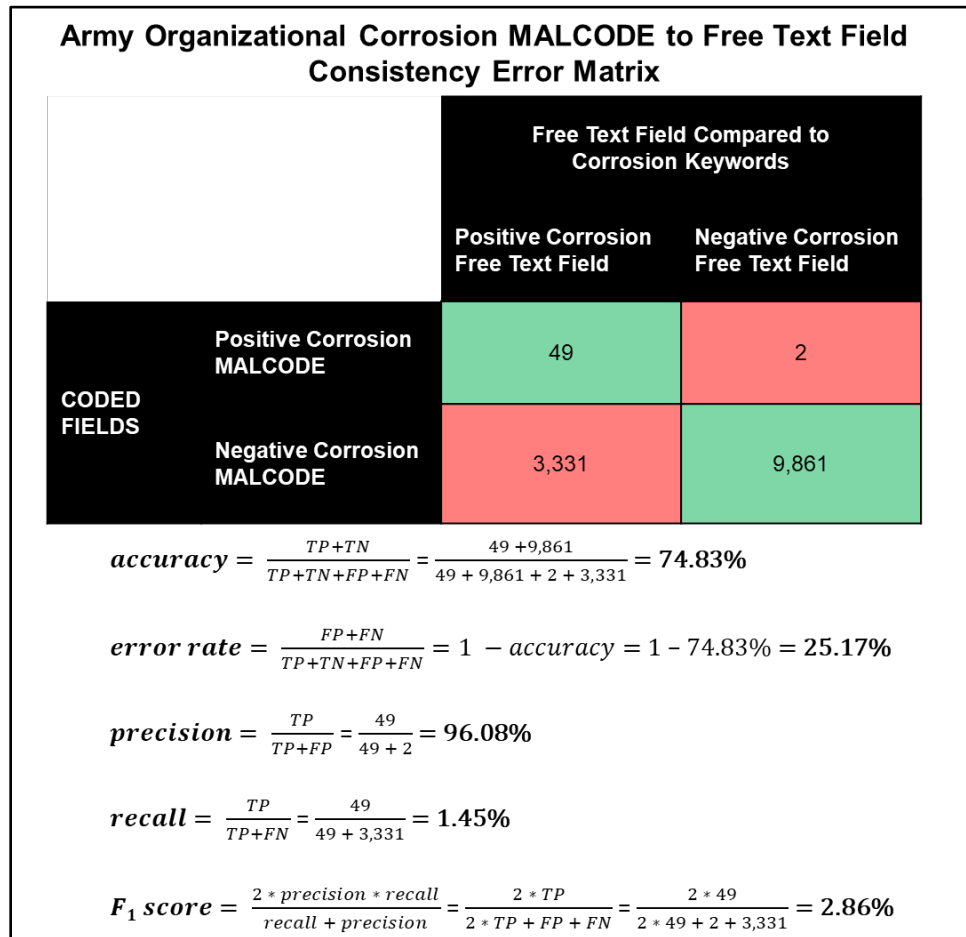


Figure 20: Army HMMWV Organizational Corrosion-Consistency Error Matrix and Performance Metrics

For Marine Corps organizational HMMWV maintenance records, 9,322 records were assessed for consistency and completeness for 'maintenance malfunction.' When comparing the Defect Code field to the free text field (Problem Summary), an estimated 35.14% of records had Jaro-Winkler similarity scores that indicated 'match.' 22.57% of records had a Jaro-Winkler similarity score of [0.92 to 1.0] (match) and 17.95% of records had a Jaro-Winkler similarity score of [0.79 to 0.92) (potential match). Applying the 0.70 match proportion to the 'potential match' classification yields the 35.14% estimated malfunction consistency match score. The 95% confidence interval for the comparison yields a range of 32.01% to 38.27%. Figure 38 highlights the distribution of 'match,' 'potential match,' and 'non-match' records. The Marine Corps organizational HMMWV maintenance data have 644 records of a possible 9,322 records with blank Defect Codes (6.91%). This represents a fairly complete maintenance record set. The very low estimated match score (35.14%) is likely related to the very succinct Problem Summary free text fields and a lack of other free text descriptions in each maintenance record.

For Marine Corps organizational HMMWV maintenance records, 9,322 records were assessed based on Defect Code to determine if a record was corrosion-related. 784 of the records had a 'corrosion-positive'



Defect Code, and 8,538 of the records had a 'corrosion-negative' Defect Code. Figure 39 depicts the distribution of 'corrosion-positive' records that had free text narratives that included corrosion keywords from the MADW Corrosion Dictionary. 60.33% of records with corrosion-positive Defect Codes had corrosion-related free text. Similarly, Figure 40 depicts the distribution of 'corrosion-negative' records that had free text narratives that included corrosion keywords from the MADW Corrosion Dictionary. 15.98% of records with corrosion-negative Failure Codes still had corrosion-related free text. Figure 41 then combines these results to develop the error matrix and performance metrics for Marine Corps organizational HMMWV maintenance data. The Marine Corps organizational maintenance records are fairly accurate (more accurate than Army HMMWV records), with an accuracy score of 82.03%. The records lack precision (60.33%), meaning that 60.33% of the time, a record with a 'corrosion-positive' Defect Code will also have free text with corrosion-related keywords (based on coded and free text field consistency). Recall is very low (25.75%), indicating that of records with corrosion-related free text, 25.75% of those records are properly classified. Finally, the F₁ score (weighted average of precision and recall that penalizes extreme values) is 36.09%, which aligns with the low precision and recall scores. Based on this analysis, it is likely that corrosion-related records do not have a corrosion-related Defect Code assigned to them. This may be because multiple Defect Codes can be assigned to an individual record, and users do not always enter all relevant Defect Codes.

Having assessed the Army and Marine Corps HMMWV organizational maintenance records a few key findings emerge related to the quality of the data. As expected, the data was lower quality than the data for H-60 platform. This is because additional systems and checks are in place for H-60 aircraft to ensure pilot and crew safety for an airborne system. When assessing the HMMWV maintenance data, it was immediately clear that records did not have the fidelity or quality of the H-60 data. The Marine Corps data did not include specific part codes/descriptions, nor did it include action taken codes/descriptions. Without these data fields, consistency and completeness assessments for HMMWV data were not possible for the Marine Corps records. There were no discernable differences when comparing UAH (M1151A1) to the NAH (M1097A2). This is likely because vehicles receive maintenance at the same locations/units and the users/maintainers are often also the same individuals. When assessing the Failure Codes for the Army HMMWV records, an unanticipated finding emerged. 70.12% of the Army records had a '099' Failure Code (OTHER/UNKNOWN/BROKEN), which is problematic when trying to assess the maintenance issue for a given maintenance record. Pairing this with the fact that 98.55% of maintenance records had 'corrosion-positive' free text but were classified as 'corrosion-negative' based on the Failure Code highlights an area of concern. The Army may want to consider two actions: improve the training for maintainers to decrease the usage of Failure Code '099,' and provide additional corrosion-related Failure Codes that capture the maintenance discrepancies that exist. Table 13 highlights the accuracy score and F₁ score for corrosion-related malfunctions when classifying the coded fields and free text fields as 'corrosion-positive' or 'corrosion-negative' based on the MADW Corrosion Dictionary. The F₁ scores for both Army and Marine Corps HMMWV records are very low. While both record types had somewhat high precision scores (96.08% and 60.33% respectively), recall scores were very low (1.45% and 25.75% respectively). This yields low F₁ scores. The key finding for both Army and Marine Corps HMMWV corrosion-related records is that a high percentage of records have 'corrosion-positive' text in the narrative description but are assigned a 'corrosion-negative' Failure/Defect Code.

Table 13: Results for HMMWV Maintenance Records

Record Type	Accuracy for Corrosion Match Scoring	F ₁ Score for Corrosion Matching
Army Organizational HMMWV Maintenance Data	74.83%	2.86%
Marine Corps Organizational HMMWV Maintenance Data	82.03%	36.09%

4.4. Findings: Recommended Ground Systems for Future Study

The methodology developed to assess maintenance data quality can be expanded to additional systems for future study. Attempting to identify additional systems for future study was a subjective task; however, maintenance SMEs as well as senior decision makers (former company, battalion, and brigade commanders) were interviewed, and their expertise weighed heavily on recommendations. The suggested ground systems for future study span several warfighting functions (movement and maneuver, fires, and mission command) [38, pp. 15-16]. Applying a similar maintenance data quality approach to a broad set of systems would highlight if data quality trends and issues are shared across a diverse set of military systems or generally isolated to rotary-wing and ground vehicle systems.

The Joint Light Tactical Vehicle (JLTV) is a vehicle family currently being fielded to replace the HMMWV for the Army and Marine Corps and supports the movement and maneuver warfighting function. It is expected to “provide better protection for passengers against current and future battlefield threats, increased payload capacity, and improved automotive performance over the up-armored HMMWV” [39, p. 97]. As the replacement for the HMWMV, the JLTV will be a heavily utilized vehicle that will be supported by the same maintainers and data systems (GCSS-A/GCSS-MC) as the HMMWV. By evaluating the JLTV’s maintenance and corrosion data quality, it can be assessed whether the poor data quality found in HMMWVs is a product of older vehicles with numerous maintenance issues, or if the issue is more systemic across ground vehicle platforms.

The M142 High Mobility Artillery Rocket System (HIMARS) is a Multiple Launch Rocket System (MLRS) used by the Army and Marine Corps and supports the fires warfighting function. HIMARS are deployed globally to provide strategic capability and improve homeland and important asset defense. HIMARS are considered extremely versatile, capable of firing the entire MLRS family of munitions [40]. Given the strategic importance of HIMARS, evaluating maintenance and corrosion data quality is worthwhile. This assessment can determine how fires systems maintenance and corrosion data compare to ground vehicles and rotary-wing platform data quality.

The AN/PRC-117 Multiband Manpack Radios/Multiband Multi Mission Radio (MBMMR) provides advanced network communication for warfighters and supports the mission command warfighting function. MBMMRs allow for wideband and narrowband connectivity, and satellite communication along with secure voice and data encryption capability (through TOP SECRET). MBMMRs are used by both Army and Marine Corps units and enable beyond line of sight communications [41]. MBMMR and communication system maintenance and corrosion data quality assessment would provide insights related to smaller, individualized systems. Comparisons can be made between maintenance data quality for the MBMMR and other warfighting systems.

Two weapon systems (M4 carbine and M2 .50 caliber machine gun) are recommended for future study related to maintenance and corrosion data quality. The M4 carbine and M2 .50 caliber machine gun support



the movement and maneuver warfighting function. The M4 carbine is classified as an individual weapon system, and the M2 .50 caliber machine gun is classified as a crew-served weapon system. Both the M4 and M2 are used in all Services (Army, Navy, Air Force, and Marine Corps). The M4 carbine is the standard issue firearm for most units in the U.S. military, and the M2 .50 caliber machine gun is a common crew-served system for ground vehicles, aircraft, and naval ships [42] [43]. Assessing the maintenance and corrosion data quality would be beneficial, and likely extremely variable since the systems are used by most military units in varying conditions. Best practices and highest quality data could be identified and recommendations to improve data collection could prove beneficial.

These recommendations (JLTV, HIMARS, AN/PRC-117, M4 carbine, and M2 .50 caliber machine gun) represent a broad set of military systems that provide a diverse set of functions across Services. Further analysis into the maintenance and corrosion data quality will yield insights related to the data collection process best practices and areas needing improvement.



Section 5

Conclusions and Implications

5.1. Summary of Findings

Having assessed H-60 and HMMWV records for consistency, completeness, and accuracy the findings are best summarized by objective.

Objective 1 – Data Quality for Navy and Air Force H-60 Maintenance Records:

1. Having assessed the Navy and Air Force H-60 maintenance records (both organizational and depot-level), no single record type is dominant when considering the multiple hypotheses that were assessed.
2. For the object receiving maintenance, the Air Force H-60 depot-level maintenance records scored the highest, with a 96.19% object match consistency score. For maintenance action taken, Navy H-60 organizational maintenance records scored the highest, with an action taken match consistency score of 84.05%.
3. For maintenance discrepancy, Navy H-60 organizational maintenance records again scored the highest, with a malfunction match consistency score of 78.52%.

When comparing corrosion-coded fields to free text fields, comparing F₁ scores would have been ideal to compare across record sets; however, with the significantly different proportions of corrosion-positive MALCODES, accuracy is the most reasonable metric to use for comparison across datasets.

1. Navy and Air Force organizational and depot-level H-60 maintenance records exhibit high accuracy levels (all exceed 83%) (see Table 14).
2. Air Force H-60 organizational maintenance records have the highest accuracy when comparing coded fields to free text fields and classifying both the coded fields and free text fields into corrosion-positive and corrosion-negative categories.

Air Force maintenance record completeness is higher than for the Navy maintenance records (fewer blanks or coded fields not in TMs). Object data fields had the lowest record completeness for both Services of all categorizes that were evaluated. This is likely because WUCs are manually entered and are five to eight characters in length and there are several thousand possible WUCs (7,399 WUCs for Navy, 2,319 WUCs for Air Force), which introduces more opportunity for input errors. Overall, the Navy and Air Force H-60 maintenance data was determined to be higher quality than the Army H-60 maintenance data assessed in the FY20 CPO Study. The Army corrosion tag accuracy was assessed to be 41.6% for ASAP-RAM H-60 maintenance data. However, as the Army transitions to Aircraft Notebook, record accuracy should improve, and the amount of free text/manual entries will decrease. The Army has already identified some of their issues with maintenance quality, hence the shift to Aircraft Notebook for data collection.

**Table 14: Navy and Air Force H-60 Corrosion Consistency and Completeness Summary**

Record Type	Corrosion Match Accuracy	Completeness
Navy Organizational H-60 Maintenance Data	83.60%	88.05% WUCs complete
Navy Depot H-60 Maintenance Data	92.66%*	No AT Codes available in records
Air Force Organizational H-60 Maintenance Data	96.57%	98.31% WUCs complete
Air Force Depot H-60 Maintenance Data	88.11%	98.66% MALCODES complete

*Navy Depot H-60 Maintenance Records are all corrosion-positive.

Objective 2 – Best Practices and Recommendations Based on H-60 Data Quality Assessment:

Having assessed the H-60 maintenance collection processes for the Navy, Air Force, and Army, a few best practices and recommendations emerge.

Best practices include establishing feedback loops within the data collection process, establishing appropriate corrosion-related codes, and having engaged unit leadership. The Navy ADDR program communicates potential deficiencies from maintenance that should have been executed at the organizational level. FRC Aircraft E&Es assess an aircraft at the depot level and provide detailed descriptions of all unscheduled organizational level discrepancies, generating an ADDR. The Close Out ADDR contains summary information about the overall cost and delivery of the aircraft for the DLM event [37]. Air Force DITs focus on ensuring squadron (organizational level) maintenance data is complete and accurate. An additional step that DITs focus on is root cause analysis for poor data quality. DITs are comprised of organizational level maintainers who “peer-review” the maintenance entries of other units. The Navy and Air Force have established appropriate corrosion-related malfunction codes. While ground vehicles in the Army and Marine Corps have very limited numbers of corrosion-related malfunction codes (<5 dedicated malfunction codes), the Air Force has 10 corrosion-specific malfunction codes, and the Navy has 49 corrosion-specific malfunction codes. This specificity is advantageous to allow maintainers to isolate corrosion to specific failure modes. A final best practice is to ensure maintenance unit leaders remain engaged. Having supervisors (enlisted, officer, and civilian) conduct frequent quality checks helps ensure attention to detail and helps train maintainers to improve their maintenance knowledge, skills, and abilities.

Recommendations focus on improving data collection at point of entry through NLP AI tools (predictive text), improving training and standardization for data collection, requiring specific mandatory data fields, and incorporating other forms of data (images, additional coded fields, etc.). A major emphasis needs to be placed on improving data collection at point of entry through AI tools (predictive text). This could include mobile applications to document maintenance actions that limit free text entry and incorporate other forms of data, such as images of the equipment receiving maintenance or voice-to-text to improve description detail. Descriptions of maintenance work done could be compiled from PMCS procedures and the equipment technical manuals such that NLP AI systems could complete the input of a user as they begin to describe the maintenance action through voice or text. This technology already exists in the civilian sector and could be adapted for use in the military.



A second key recommendation is to improve user/maintainer training and standardize data collection. Different units and locations have vastly different standards for what is an acceptable maintenance free text description, how detailed a malfunction code or WUC should be, and what shorthand/acronym is included in a description. When each organization has its own language, deciphering at the Service or OSD level is challenging. By standardizing what is acceptable for a free text description, better insights can be drawn from maintenance records. Another issue that training needs to address is increasing the specificity of maintenance records. WUCs and MALCODES should be as specific as possible, but often generic codes are used for convenience or lack of user knowledge.

A final recommendation for data quality relates to ensuring data is complete and interpretable. Maintenance forms should have some specific (critical) mandatory fields. This would eliminate some of the completeness issues identified within the maintenance records. Additionally, drop-down/autofill features for completing maintenance forms could ensure that coded fields align with current technical manuals and regulations.

Objective 3 – Data Quality for Army and Marine Corps HMMWV Maintenance Records:

Having assessed the Army and Marine Corps HMMWV organizational maintenance records a few key findings emerge related to the quality of the data. As expected, the data was lower quality than the data for H-60 platform. This is because additional systems and checks are in place for H-60 aircraft to ensure pilot and crew safety for an airborne system. When assessing the HMMWV maintenance data, it was immediately clear that records did not have the fidelity or quality of the H-60 data.

1. There were no discernable differences when comparing UAH (M1151A1) to the NAH (M1097A2). This is likely because vehicles receive maintenance at the same locations/units and the users/maintainers are often also the same individuals.
2. When assessing the Failure Codes for the Army HMMWV records, an unanticipated finding emerged. 70.12% of the Army records had a '099' Failure Code (OTHER/UNKNOWN/BROKEN), which is problematic when trying to assess the maintenance issue for a given maintenance record. Pairing this with the fact that 98.55% of maintenance records had 'corrosion-positive' free text but were classified as 'corrosion-negative' based on the Failure Code highlights an area of concern.
3. The Army may want to consider two actions: improve the training for maintainers to decrease the usage of Failure Code '099,' and provide additional corrosion-related Failure Codes that capture the maintenance discrepancies that exist.
4. The Marine Corps HMMWV records had 74.25% of records that had corrosion-positive free text but were classified with a corrosion-negative Defect Code. Table 15 highlights the corrosion accuracy and completeness for Army and Marine Corps organizational level HMMWV maintenance records.

The key finding for both Army and Marine Corps HMMWV corrosion-related records is that a high percentage of records have 'corrosion-positive' text in the narrative description but are assigned a 'corrosion-negative' Failure/Defect Code.

**Table 15: Army and Marine Corps HMMWV Corrosion Consistency and Completeness Summary**

Record Type	Corrosion Match Accuracy	Completeness
Army Organizational HMMWV Maintenance Data	74.83%	99.06% malfunctions complete
Marine Corps Organizational HMMWV Maintenance Data	82.03%	No Part Codes or AT Codes available in records

Objective 4 – Recommended Ground Systems for Future Study:

Maintenance SMEs as well as senior decision makers (former battalion and brigade commanders) were interviewed, and their expertise weighed heavily on recommendations. The suggested ground systems for future study span several warfighting functions (movement and maneuver, fires, and mission command) [35, pp. 15-16]. Applying a similar maintenance data quality approach to a broad set of systems would highlight if data quality trends and issues are shared across a diverse set of military systems or generally isolated to rotary-wing and ground vehicle systems.

- Recommended Systems for Future Study:
 - Joint Light Tactical Vehicle (JLTV)
 - High Mobility Artillery Rocket System (HIMARS)
 - Multiband Multi Mission Radio (MBMMR)
 - M4 Carbine
 - M2 .50 Caliber Machine Gun

These recommendations (JLTV, HIMARS, AN/PRC-117, M4 carbine, and M2 .50 caliber machine gun) represent a broad set of military systems that provide a diverse set of functions across Services. Further analysis into the maintenance and corrosion data quality will yield insights related to the data collection process best practices and areas needing improvement.

5.2. Impact of the Study

This study highlights that Navy and Air Force H-60 maintenance records are of high quality and it is reasonable to use this data to make decisions. MADW algorithms can further clean and process the raw maintenance data from the Navy and Air Force H-60 platforms to gain insights about corrosion-related issues; however, having good quality data is a pre-requisite for more in-depth analysis. Additionally, there should be a high level of confidence in the H-60 coded fields for the objects receiving maintenance, the action taken, and the malfunction based on comparisons between the coded and free text data fields. The best practices and recommendations from the H-60 maintenance data analysis highlights the need for feedback loops, engaged small unit leaders, AI tools at the point of entry, additional training and standardization for maintainers, and mandatory fields to reduce missing information.

The second key area to highlight is that ground systems may have significantly lower maintenance data quality than aviation systems, based on the analysis of Army and Marine Corps HMMWV maintenance



records. While the overall accuracy of HMMWV maintenance data appears almost comparable to the H-60 accuracy scores, a significant percentage of records with corrosion-related words are not tagged with a corrosion-related malfunction code. For the Army HMMWV records, this occurred 98.55% of the time, while for the Marine Corps this occurred 74.75% of the time. Both results are cause for concern, especially considering that the worst similar rate in H-60 records was 41.88%. Ground system maintenance quality requires significant improvement to become close to the level of Navy and Air Force H-60 maintenance records. This indicates a need for further standardization and training across the maintenance workforce.

5.3. Recommended Future Work

The low precision, recall, and F_1 scores calculated in the evaluation of maintenance records (both H-60 and HMMWV) paired with the consistency scores for maintenance object, maintenance action, and maintenance malfunction suggest that future efforts should be put into improving the data collection and data cleaning processes. Because maintenance data is collected at the lowest levels across the Services, standardization of data collection methods is very important to reduce the variability in maintenance records. A set of common data collection tools and processes would be very helpful to allow users of varying knowledge, skills, and abilities to contribute meaningful and accurate data entries into the maintenance system.

Future work should include assessing the recommended ground systems identified in Objective 4 to determine a baseline data quality for systems beyond the H-60 and HMMWV platforms. Incorporating AI tools at the point of maintenance and data collection would allow for cleaner data for MADW to process. This would provide more meaningful analysis to senior decision makers. Mobile applications could be used to document maintenance actions and limit free text entry while incorporating other forms of data, such as images and voice-to-text. Descriptions of maintenance work done could be compiled from PMCS procedures and equipment technical manuals such that NLP AI systems could complete the input of a user as they begin to describe the maintenance action through voice or text (predictive sentence completion). This technology already exists in the civilian sector and could be adapted for use in the military.

Another area for future work is to conduct additional analysis on a broader set of H-60 and HMMWV maintenance records. Given the COVID-19 pandemic and the constraints it created on the ability to travel and meet key stakeholders in the maintenance process, access to data was a significant challenge for this study. Gaining access to additional records from the Services would confirm this study's findings. Rather than performing the analysis on a sample of maintenance records from each Service for the respective system, using multiple years of maintenance records would provide more definitive insights.



Appendix A References

- [1] United States Government Accountability Office, "Observations on Changes to the Reporting Structure for DOD's Corrosion Office and Its Implementation of GAO Recommendations," Government Accountability Office, Washington, D.C., 2019.
- [2] United States Government Accountability Office, "Opportunities to Reduce Corrosion Costs and Increase Readiness," Government Accountability Office, Washington, D.C., 2003.
- [3] Defense Acquisition University, "Corrosion Prevention and Control," 2019. [Online]. Available: <https://www.dau.edu/acquipedia/pages/articledetails.aspx#!507>. [Accessed 24 February 2021].
- [4] Office of the Assistant Secretary of Defense for Sustainment, "Corrosion Prevention," 2020. [Online]. Available: https://www.acq.osd.mil/log/MR/corrosion_prev.html. [Accessed 24 February 2021].
- [5] United States Government Accountability Office, "DOD Should Take Additional Actions to Enhance Corrosion Prevention and Mitigation Efforts," Government Accountability Office, Washington, D.C., 2018.
- [6] Office of the Assistant Secretary of Defense for Sustainment, "Maintenance and Availability Data Warehouse (MADW)," 2021. [Online]. Available: <https://www.acq.osd.mil/log/mpp/MADW.html>. [Accessed 25 February 2021].
- [7] Logistics Management Institute (LMI), *MADW Virtual User Conference*, DoD MADW, 23 February 2021.
- [8] Logistics Management Institute (LMI), *2019 Maintenance and Availability Data Warehouse (MADW) - CCPE and CPO Briefing*, DoD MADW, August 2019.
- [9] Logistics Management Institute (LMI), "DOD Sustainment Dashboard," 25 February 2021. [Online]. Available: <https://madw.acq.osd.mil/dashboards/DOD%20Sustainment>. [Accessed 25 February 2021].
- [10] J. Case, "USMA 2020 Corrosion Data Analysis Study: Estimating MADW Object Identification Accuracy and Corrosino Algorithm Improvement," United States Military Academy: Operations Research Center, West Point, NY, 15 June 2020.
- [11] Department of the Army, "The Army Maintenance Management Systems (TAMMS) User Manual," Headquarters, Department of the Army, Washington, D.C., 22 August 2005.
- [12] Department of the Navy, "COMNAVAIRFORINST 4790.2C: The Naval Aviation Maintenance Program," Headquarters Department of the Navy, Washington, D.C., 15 January 2017.
- [13] Department of the Navy, "MCTP 3-40E: Maintenance Operations," Headquarters, United States Marine Corps, Washington, D.C., 28 January 2020.
- [14] Department of the Air Force, "Air Force Instruction 21-101: Aircraft and Equipment Maintenance Management," Headquarters Department of the Air Force, Washington, D.C., 15 September 2020.

-
- [15] Department of the Air Force, "TO 00-20-2: Technical Manual - Maintenance Data Documentation," Headquarters Department of the Air Force, Washington, D.C., 15 March 2016.
- [16] F. Sidi, P. H. S. Panahy, L. S. Affendey, M. A. Jabar, H. Ibrahim and A. Mustapha, "Data Quality: A Survey of Data Quality Dimensions," in *2012 International Conference on Information Retrieval & Knowledge Management*, Kuala Lumpur, Malaysia, 2012.
- [17] R. Y. Wang and D. M. Strong, "Beyond Accuracy: What Data Quality Means to Data Consumers," *Journal of Management Information Systems*, vol. 12, no. 4, pp. 5-33, 1996.
- [18] C. Batini, C. Cappiello, C. Francalanci and A. Maurino, "Methodologies for Data Quality Assessment and Improvement," *ACM Computing Surveys*, vol. 41, no. 3, pp. Article 16, pages 1-52, 2009.
- [19] B. Lantz, *Machine Learning with R: Expert Techniques for Predictive Modeling*, Birmingham: Packt Publishing, 2019.
- [20] D. Vatsalan, P. Christen and V. Verykios, "A Taxonomy of Privacy-Preserving Record Linkage Techniques," *Information Systems*, vol. 38, no. 6, pp. 946-969, 2013.
- [21] G. Navarro, "A Guided Tour to Approximate String Matching," *ACM Computing Surveys*, vol. 33, no. 1, pp. 31-88, 2001.
- [22] M. Jaro, "UNIMATCH: A Record Linkage System - Users Manual," U.S. Bureau of the Census, Washington, D.C., 1978.
- [23] I. Kloo, M. F. Dabkowski and S. H. Huddleston, "Improving Record Linkage for Counter-Threat Finance Intelligence with Dynamic Jaro-Winkler Thresholds," in *2019 Winter Simulation Conference*, National Harbor, MD, 2019.
- [24] W. E. Winkler, "Overview of Record Linkage and Current Research Directions," U.S. Census Bureau, Washington, D.C., 2006.
- [25] Naval Air Systems Command, "H-60 Multi-Mission Helicopters Program," 9 April 2021. [Online]. Available: <https://www.navair.navy.mil/organization/PMA-299>.
- [26] Y. Sverdlik, "Big Data Gives Real-Time Logistics to US Navy," 7 February 2012. [Online]. Available: <http://archive.datacenterdynamics.com/focus/archive/2012/02/big-data-gives-real-time-logistics-us-navy#:~:text=Deckplate%20data%20subjects%20include%20flight,system%20provides%20enterprise%2Dwide%20visibility..> [Accessed 9 April 2021].
- [27] Air Force Corrosion Prevention and Control Office, "AFCPCO - About Us," [Online]. Available: <http://www.afcpo.com/about-us.html>. [Accessed 12 April 2021].
- [28] Mercer Engineering Research Center, "MERC Engineering Services," Mercer University, [Online]. Available: <https://www.merc-mercer.org/services/>. [Accessed 12 April 2021].
- [29] Department of the Air Force, "TO 1H-60(H)G-06: Organizational Maintenance - Work Unit Code Manual for USAF HH-60G Aircraft," Headquarters Department of the Air Force, Washington, D.C., 15 August 2010.
- [30] J. Eisenstein, *Introduction to Natural Language Processing*, Cambridge, MA: The MIT Press, 2019.



-
- [31] S. Ghosh and D. Gunning, *Natural Language Processing Fundamentals*, Birmingham: Packt Publishing, 2019.
- [32] S. Vajjala, B. Majumder, A. Gupta and H. Surana, *Practical Natural Language Processing*, Boston: MA: O'Reilly, 2020.
- [33] J. Silge and D. Robinson, *Text Mining with R: A Tidy Approach*, Boston, MA: O'Reilly, 2017.
- [34] E. Mayor, *Learning Predictive Analytics with R: Get to grips with key data visualization and predictive analytic skills using R*, Birmingham: Packt Publishing, 2015.
- [35] C. Mehta, "CBM+Journey Continues with Auto-DIT," *USAF Reliability and Maintainability*, pp. 3-4, April 2017.
- [36] R. S. Cohen, "USAF Tripling Data-Driven Maintenance Efforts in 2020," *Air Force Magazine*, 26 May 2020. [Online]. Available: <https://www.airforcemag.com/usaf-tripling-data-driven-maintenance-efforts-in-2020/>. [Accessed 30 April 2021].
- [37] Department of the Navy, "Standard Work Package: Aircraft Delivery Deficiency Report (ADDR)," NAVAIR, Washington, D.C., September 2019.
- [38] Department of the Army, "ADP 3-0: Operations," Headquarters, Department of the Army, Washington, D.C., October 2017.
- [39] United States Government Accountability Office, "Defense Acquisitions: Assessments of Selected Weapons Programs," Government Accountability Office, Washington, D.C., March 2015.
- [40] Lockheed Martin, "HIMARS: Long-range, mobile, precision fire launcher interoperable with C2 systems," Lockheed Martin Corporation, 2021. [Online]. Available: <https://www.lockheedmartin.com/en-us/products/high-mobility-artillery-rocket-system.html>. [Accessed 29 April 2021].
- [41] L3 Harris, "L3 Harris Falcon III AN/PRC-117G(V)1(C) - Multiband Networking Manpack Radio," L3 Harris, 2019. [Online]. Available: https://www.l3harris.com/sites/default/files/2020-07/cs_tc_datasheet_an-prc-117g-multiband-networking-manpack-radio.pdf. [Accessed 29 April 2021].
- [42] Military.com, "M4 Carbine," Military.com, 2021. [Online]. Available: <https://www.military.com/equipment/m4-carbine>. [Accessed 29 April 2021].
- [43] Military.com, "M2 .50 Caliber Machine Gun," Military.com, 2021. [Online]. Available: <https://www.military.com/equipment/m2-50-caliber-machine-gun>. [Accessed 29 April 2021].



Appendix B

Code Used to Assess H-60 and HMMWV Maintenance Data

The code used in this analysis is available on the author's private GitHub repository. Contact MAJ Raymond Vetter at raymond.t.vetter.mil@mail.mil to request access to the code.



Appendix C

MADW-provided Corrosion Search Terms

Table 16: MADW-provided Sample of Navy/Marine Corps Aviation Corrosion Search Terms

SearchTerm	ContainsOrLike	CorroKeyword	GrossOrder
"Aerodynamic Smoothing Compound**"	Contains	AERODYNAMIC SMOOTHING COMPOUND	1
"Critical pitting potential**"	Contains	CRITICAL PITTING POTENTIAL	2
"Oxygen concentration cell**"	Contains	OXYGEN CONCENTRATION CELL	3
"Gallionella ferruginea**"	Contains	GALLIONELLA FERRUGINEA	4
"Open-circuit potential**"	Contains	OPEN-CIRCUIT POTENTIAL	5
"Semipermeable membrane**"	Contains	MEMBRANE	6
"Caustic embrittlement**"	Contains	CAUSTIC EMBRITTLEMENT	7
"Sulfate-reducing bact**"	Contains	SULFIDE	8
"surface active agent**"	Contains	SURFACE ACTIVE AGENT	9
"Breakdown potential**"	Contains	BREAKDOWN POTENTIAL	10
"Environmental crack**"	Contains	ENVIRONMENTAL CRACK	11
"Passive-active cell**"	Contains	PASSIVE-ACTIVE CELL	12
"Mechanical binding**"	Contains	MECHANICAL BINDING	13
"sodium bicarbonate**"	Contains	SODIUM BICARBONATE	14
"Aliphatic Naphtha**"	Contains	ALIPHATIC NAPHTHA	15

Table 17: MADW-provided Sample of Air Force Corrosion Search Terms

SearchTerm	ContainsOrLike	CorroKeyword	GrossOrder
"Electrolytic cell**"	Contains	ELECTROLYTIC CELL	16
"incubation period**"	Contains	INCUBATION PERIOD	17
"long-line current**"	Contains	LONG-LINE CURRENT	18
"Beyond tolerance**"	Contains	BEYOND TOLERANCE	19
"Caustic cracking**"	Contains	CAUSTIC CRACKING	20
"hydraulic cement**"	Contains	HYDRAULIC	21
"hydrogen blister**"	Contains	HYDROGEN BLISTER	22
"induced cracking**"	Contains	INDUCED CRACKING	23
"intercrystalline**"	Contains	INTERCRYSTALLINE	24
"knifeline attack**"	Contains	KNIFELINE ATTACK	25
"PEELINGXXXREQUEST"	Contains	PEEL	26
"Radiation damage**"	Contains	RADIATION DAMAGE	27
"threshold stress**"	Contains	THRESHOLD STRESS	28
"(AUTO,ERC:A)CLAS"	Contains	LEAK	29
"(AUTO/ERC:A)CLAS"	Contains	LEAK	30

**Table 18: MADW-provided Sample of Army Aviation Corrosion Search Terms**

SearchTerm	ContainsOrLike	Placeholder	GrossOrder
strip* and not "nickel strip*" and not "nickle strip*" and not "screw strip*" and not "strip screw*" and not "weather strip*" and not "weatherstrip*" and not "rubber strip"	Contains	X	1
MASK* and not "o2 mask*" and not "oxy mask*" and not "gas mask*" and not "smoke mask"	Contains	X	2
wash* and not "washer*" and not "washington"	Contains	X	3
CL/PREP/PT/FINAL EACH* AND NOT "FOD"	Contains	X	4
CL/PREP/PT/FINAL EA* AND NOT "FOD"	Contains	X	5
CHECHSHEETXXXREQUEST AND NOT "FOD"	Contains	X	6
Aerodynamic Smoothing Compound*	Contains	X	7
Copper accelerated salt spray*	Contains	X	8
standard electrode potential*	Contains	X	9
Critical pitting potential*	Contains	X	10
Chemical vapor deposition*	Contains	X	11
Oxygen concentration cell*	Contains	X	12
"CHEEEEEEEJCK" AND NOT "FOD"	Contains	X	13
Chemical conversion coat*	Contains	X	14
Alkyl benzene sulfonate*	Contains	X	15

Table 19: MADW-provided Sample of Army/Marine Corps Ground Systems Corrosion Search Terms

SearchTerm	ContainsOrLike	CorroKeyword	GrossOrder
lk and not "-09" and not "-10" and not "-08"	Contains	LEAK	1
"Aerodynamic Smoothing Compound"	Contains	AERODYNAMIC SMOOTHING COMPOUND	2
"Critical pitting potential"	Contains	CRITICAL PITTING POTENTIAL	3
"Oxygen concentration cell"	Contains	OXYGEN CONCENTRATION CELL	4
"Gallionella ferruginea"	Contains	GALLIONELLA FERRUGINEA	5
"Open-circuit potential"	Contains	OPEN-CIRCUIT POTENTIAL	6
"Semipermeable membrane"	Contains	MEMBRANE	7
"Caustic embrittlement"	Contains	CAUSTIC EMBRITTLEMENT	8
"Open-circuit potential"	Contains	OPEN-CIRCUIT POTENTIAL	9
"Sulfate-reducing bact"	Contains	SULFIDE	10



"surface active agent**"	Contains	SURFACE ACTIVE AGENT	11
"Breakdown potential**"	Contains	BREAKDOWN POTENTIAL	12
"Environmental crack**"	Contains	ENVIRONMENTAL CRACK	13
"Passive-active cell**"	Contains	PASSIVE-ACTIVE CELL	14
"CORRISIONXXXREQUEST"	Contains	CORROSION	15

Table 20: MADW-provided Exclusionary Terms

SearchTerm	ContainsOrLike	CorroKeyword	GrossOrder
corrosion free	Contains	EXCLUDE-CORROSION FREE	1
not corrosion	Contains	EXCLUDE-NOT CORROSION	2
no corrosion	Contains	EXCLUDE-NO CORROSION	3
Contact/connection defective*	Contains	EXCLUDE-CONTACT/CONNECTION DEFECTIVE	4
Insulation breakdown*	Contains	EXCLUDE-INSULATION BREAKDOWN	5
Teeth broken on gear*	Contains	EXCLUDE-TEETH BROKEN ON GEAR	6
Brittle fracture*	Contains	EXCLUDE-BRITTLE	7
Internal failure*	Contains	EXCLUDE-INTERNAL FAILURE	8
Contamination*	Contains	EXCLUDE-CONTAMINATE	9
Corona effect*	Contains	EXCLUDE-CORONA EFFECT	10
contaminants*	Contains	EXCLUDE-CONTAMINATE	11
Crystallized*	Contains	EXCLUDE-CRYSTALLIZED	12
Intermittent*	Contains	EXCLUDE-INTERMITTENT	13
insulation*	Contains	EXCLUDE-INSULATION	14
Punctured*	Contains	EXCLUDE-PUNCTURE	15
embrittled*	Contains	EXCLUDE-EMBRITTLE	16
Brittle*	Contains	EXCLUDE-BRITTLE	17
Chipped*	Contains	EXCLUDE-CHIPPED	18
Burred*	Contains	EXCLUDE-BURRED	19
CONTAM	Contains	EXCLUDE-CONTAMINATE	20



Appendix D

Navy Corrosion-Related MALCODES

MALCODES	MAL_Descriptions
Surface Corrosion	
C01	Minor surface corrosion
C02	Moderate surface corrosion
C03	Severe surface corrosion
C04	Flight critical surface corrosion
Pitting Corrosion	
C11	Minor pitting corrosion
C12	Moderate pitting corrosion
C13	Severe pitting corrosion
C14	Flight critical pitting corrosion
Exfoliation Corrosion	
C21	Minor exfoliation corrosion
C22	Moderate exfoliation corrosion
C23	Severe exfoliation corrosion
C24	Flight critical exfoliation corrosion
Galvanic Corrosion	
C31	Minor galvanic corrosion
C32	Moderate galvanic corrosion
C33	Severe galvanic corrosion
C34	Flight critical galvanic corrosion
Intergranular Corrosion	
C41	Minor intergranular corrosion
C42	Moderate intergranular corrosion
C43	Severe intergranular corrosion
C44	Flight critical intergranular corrosion
Stress Corrosion	
C51	Minor stress corrosion
C52	Moderate stress corrosion
C53	Severe stress corrosion
C54	Flight critical stress corrosion
Fatigue Corrosion	
C61	Minor fatigue corrosion
C62	Moderate fatigue corrosion

C63	Severe fatigue corrosion
C64	Flight critical fatigue corrosion
Fretting Corrosion	
C71	Minor fretting corrosion
C72	Moderate fretting corrosion
C73	Severe fretting corrosion
C74	Flight critical fretting corrosion
Filiform Corrosion	
C81	Minor filiform corrosion
C82	Moderate filiform corrosion
C83	Severe filiform corrosion
C84	Flight critical filiform corrosion
Micro-biological Corrosion	
C91	Minor micro-biological corrosion
C92	Moderate micro-biological corrosion
C93	Severe micro-biological corrosion
C94	Flight critical micro-biological corrosion
Concentration cell Corrosion	
CA1	Minor concentration cell corrosion
CA2	Moderate concentration cell corrosion
CA3	Severe concentration cell corrosion
CA4	Flight critical concentration cell corrosion
Circuit Breakers/Relays	
W12	Corroded relay/hardware
Connectors	
W13	Corroded connector/backshell (external)
Terminal Boards/Modules/Points	
W45	Corroded terminals, posts, etc.
Functional Failure Group	
W54	Connectors - corroded (internal)
Reason for Removal Group	
170	CORRODED



Appendix E

Air Force Corrosion-Related MALCODES

MALCODES	MAL_Descriptions
170	Corroded Mild/Moderate
667	Corroded Severe
211	Corroded Internal Surfaces
212	Corroded External Surfaces
143	Damaged/Cracked Compressor Case
277	Fuel Nozzle/Oil Line Coking
279	Spray Pattern Defective
410	Lack of/or Improper Lubrication
425	Pitted, Nicked, Chipped, Scored or Scratched
846	Delaminated

Appendix F

H-60 Coded Field to Free Text Comparison

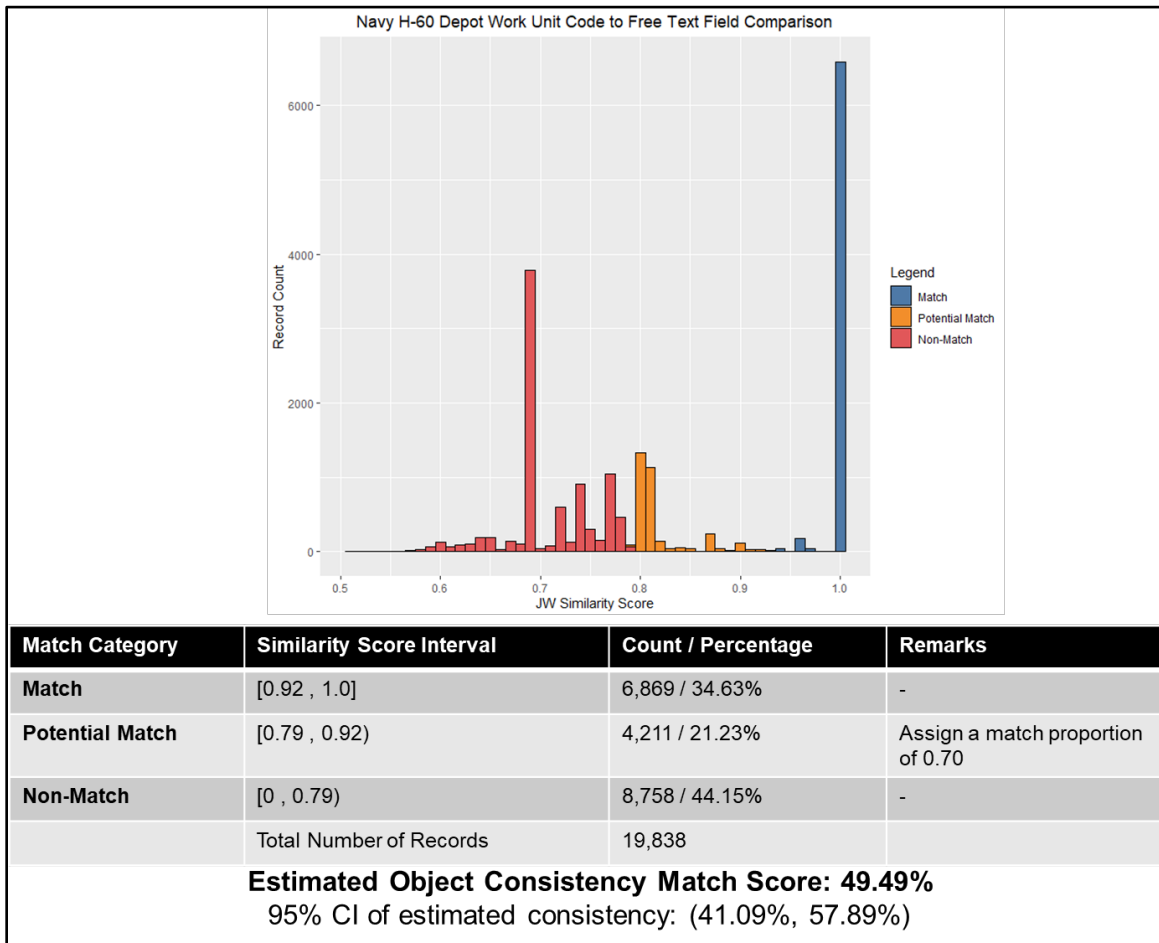


Figure 21: Navy H-60 Depot Maintenance Data Part Code to Free Text Comparison

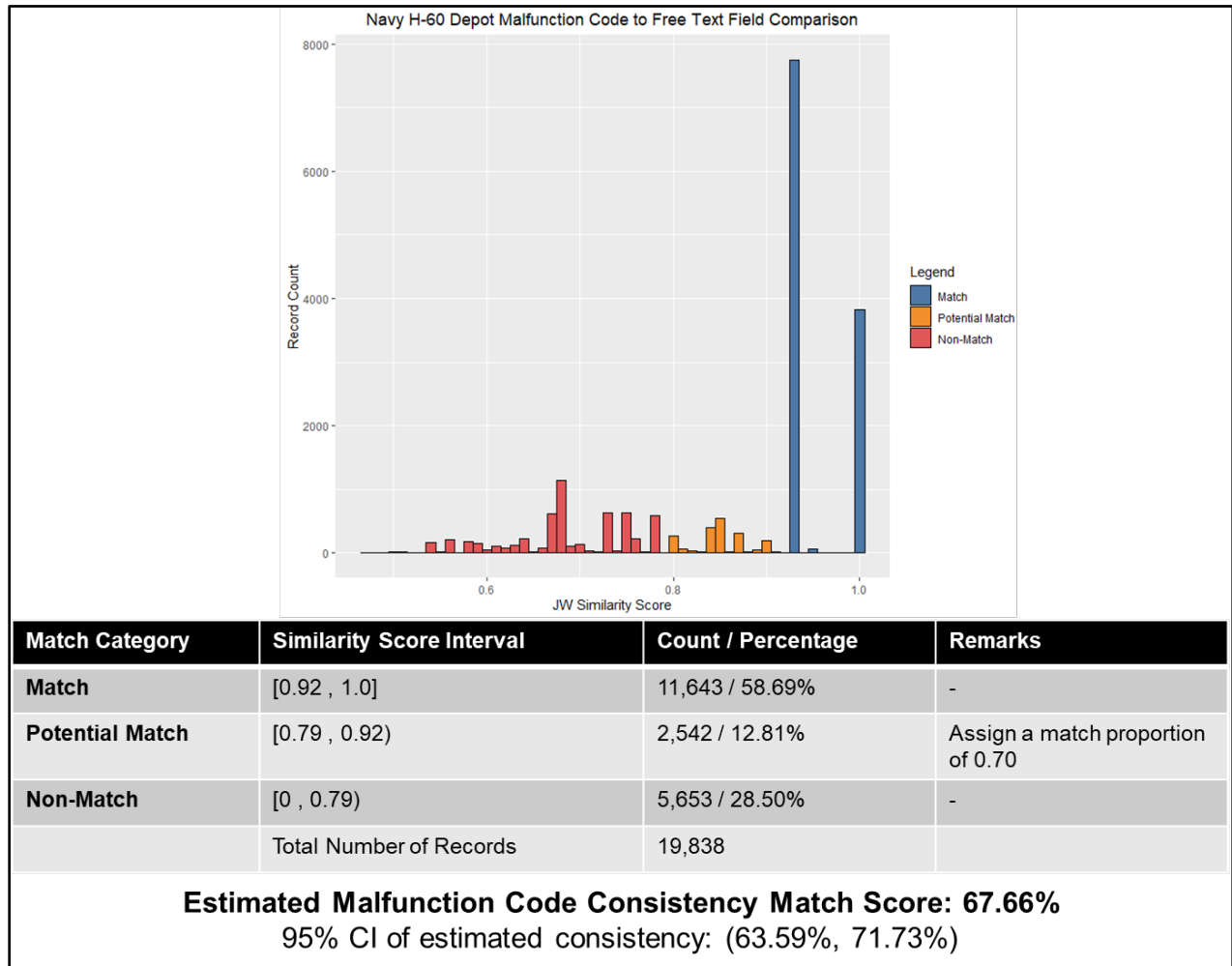


Figure 22: Navy H-60 Depot Maintenance Data Malfunction Code to Free Text Comparison

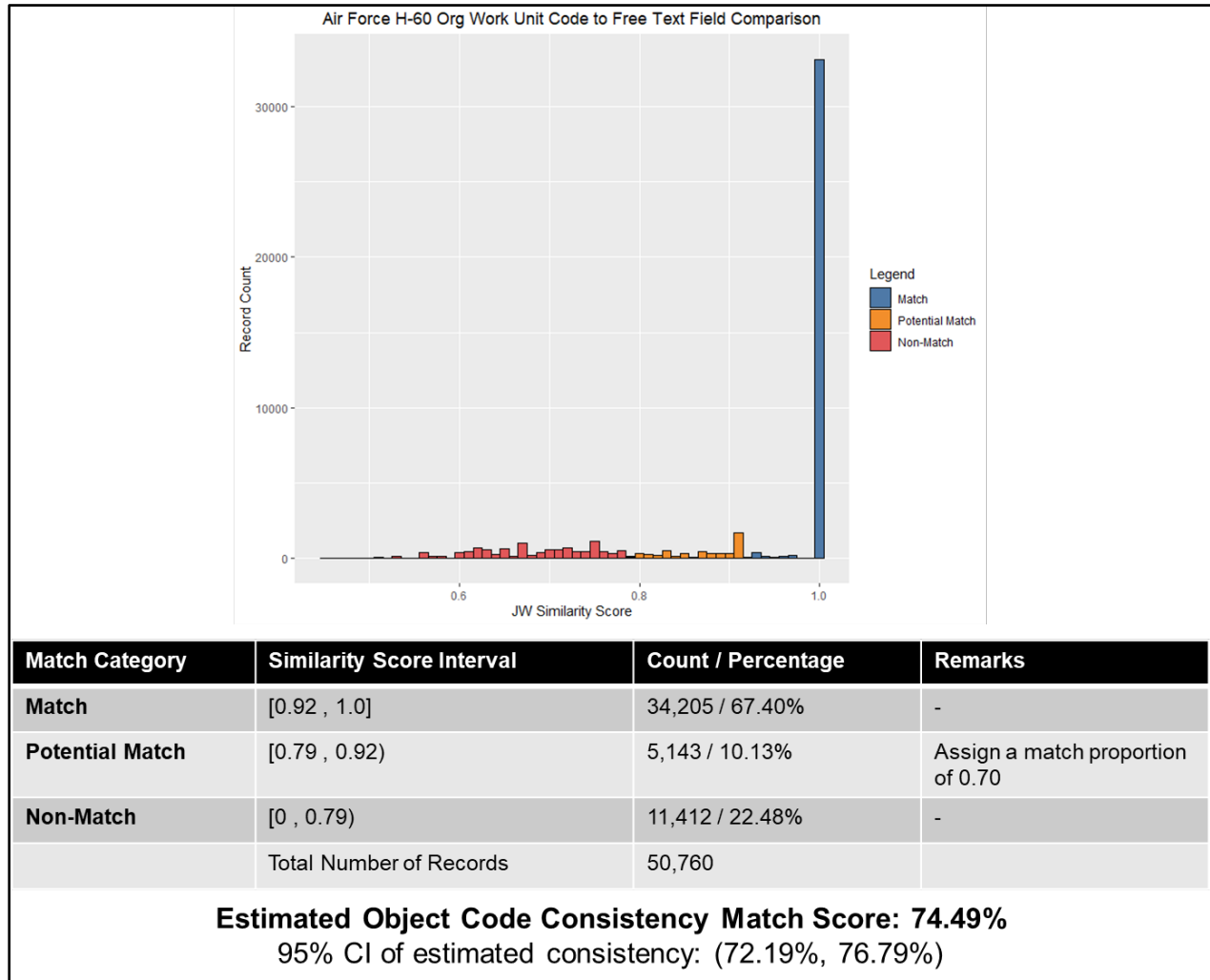


Figure 23: Air Force H-60 Organizational Maintenance Data Part Code to Free Text Comparison

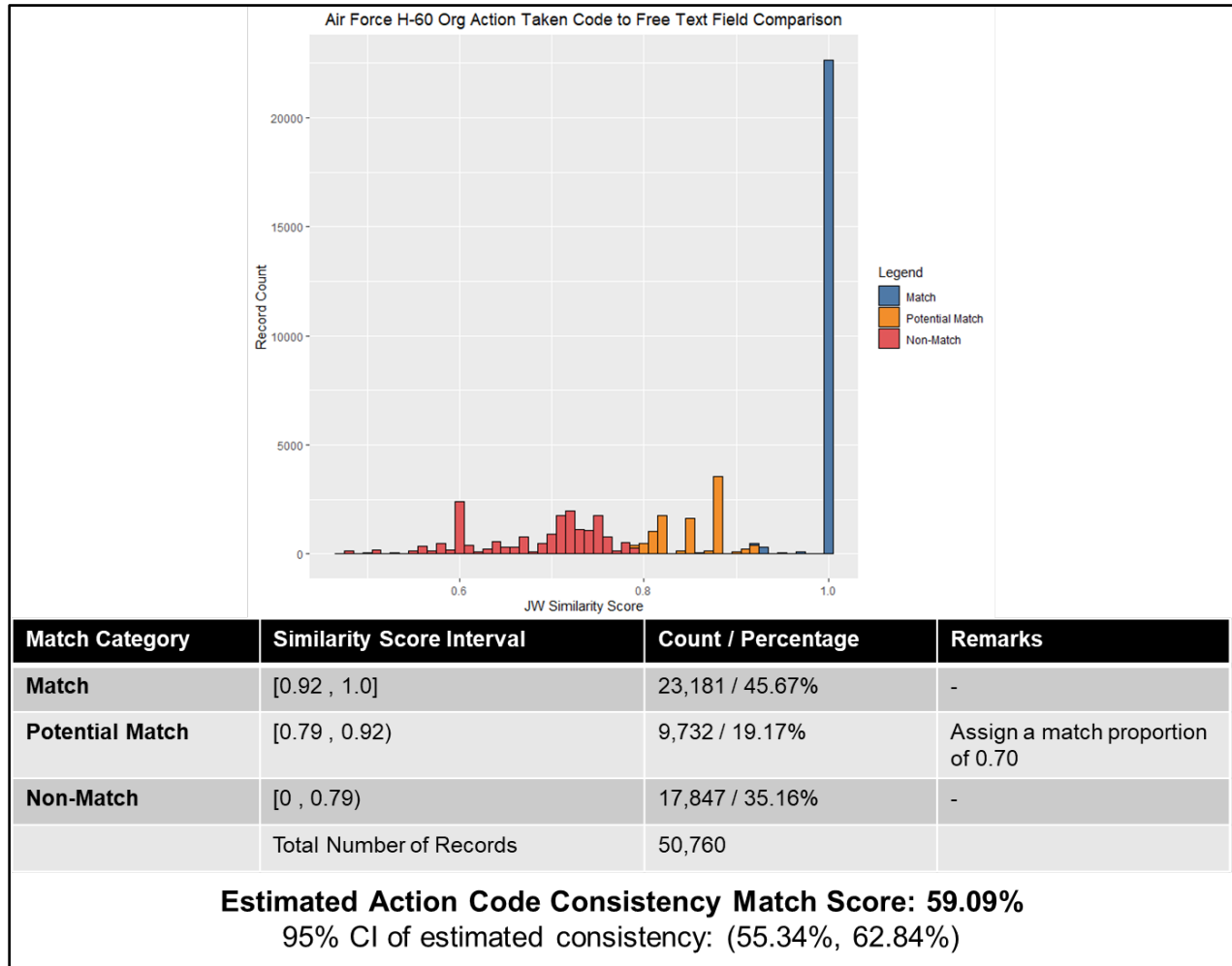


Figure 24: Air Force H-60 Organizational Maintenance Data Action Code to Free Text Comparison

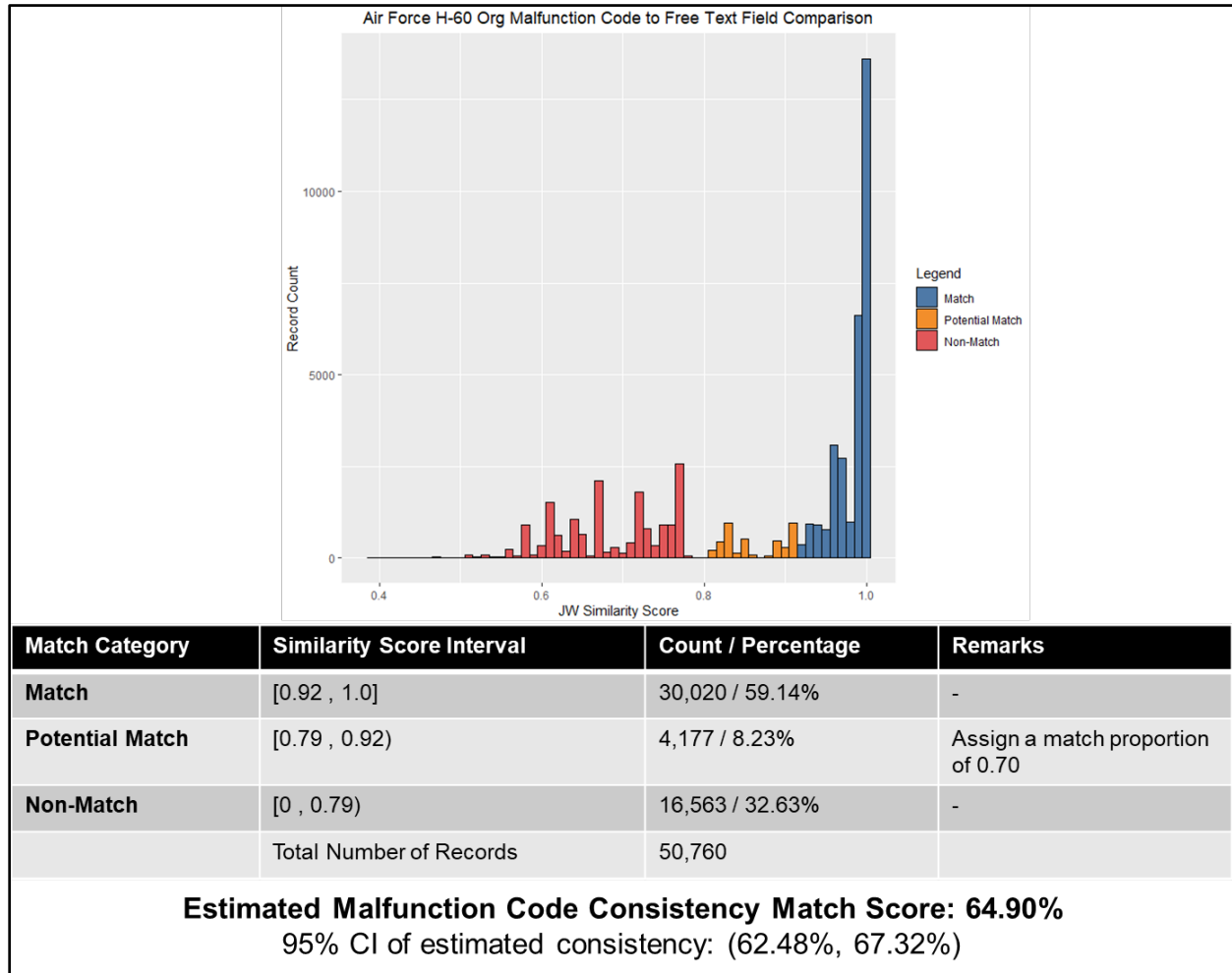


Figure 25: Air Force H-60 Organizational Maintenance Data Malfunction Code to Free Text Comparison

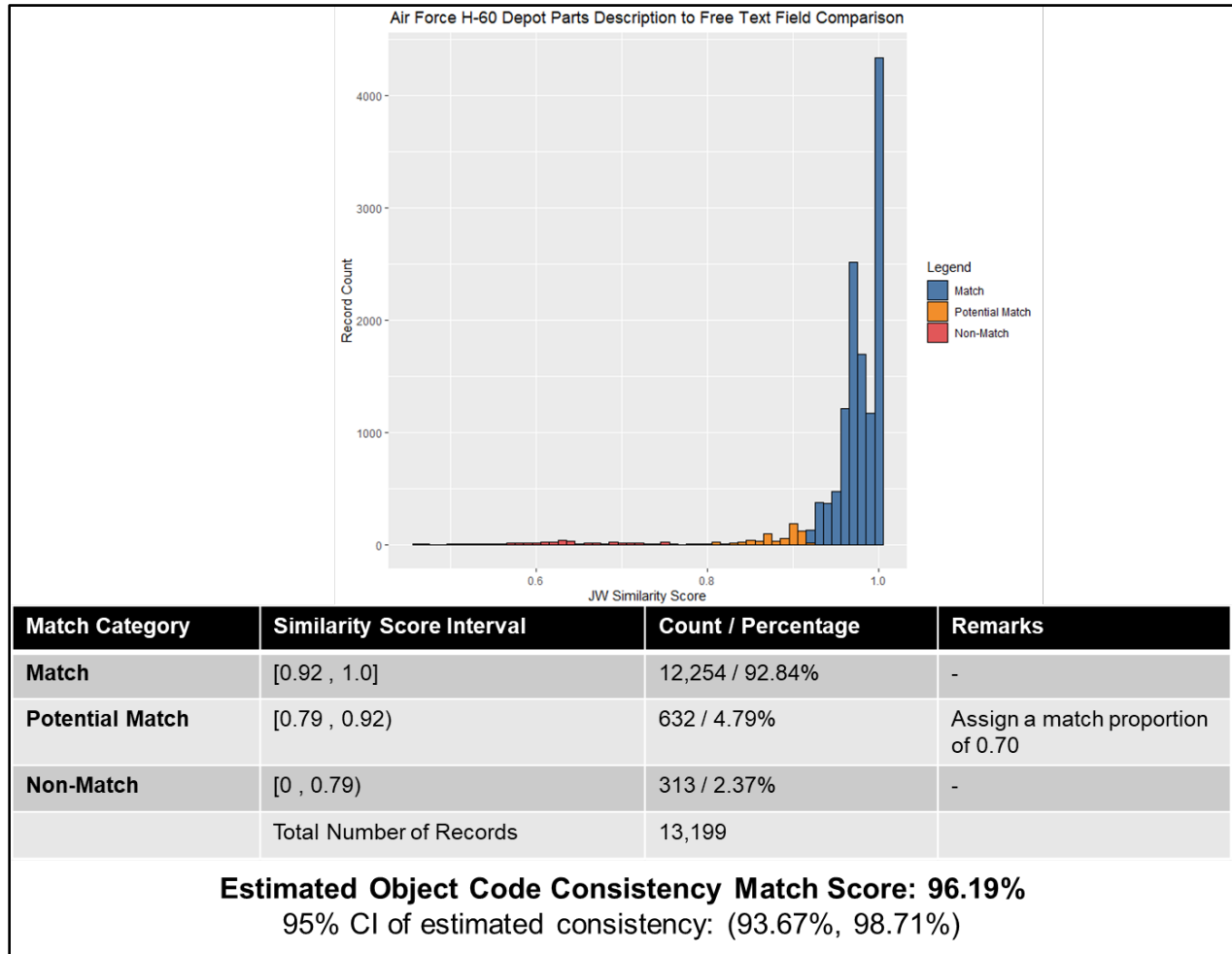


Figure 26: Air Force H-60 Depot Maintenance Data Part Code to Free Text Comparison

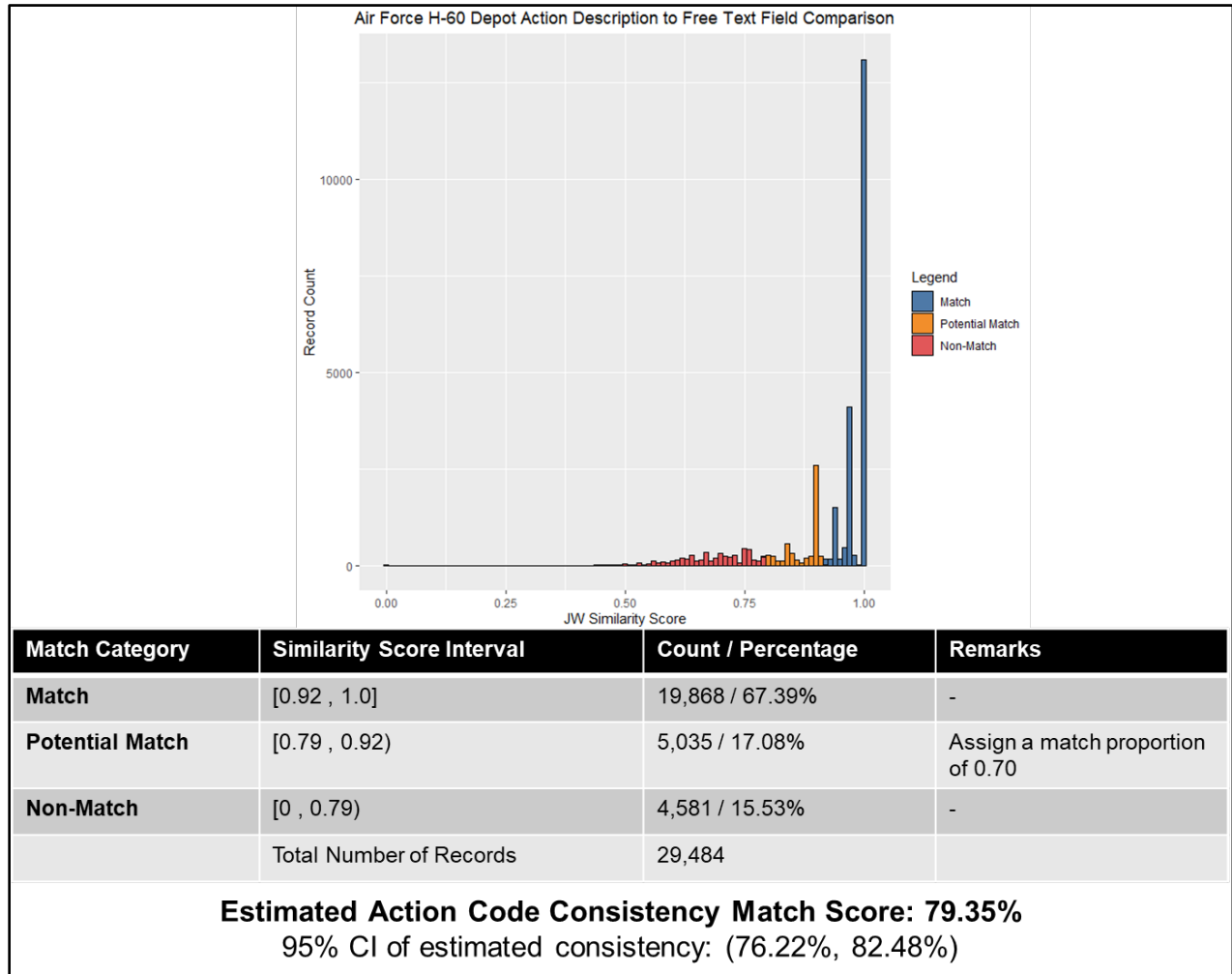


Figure 27: Air Force H-60 Depot Maintenance Data Action Code to Free Text Comparison

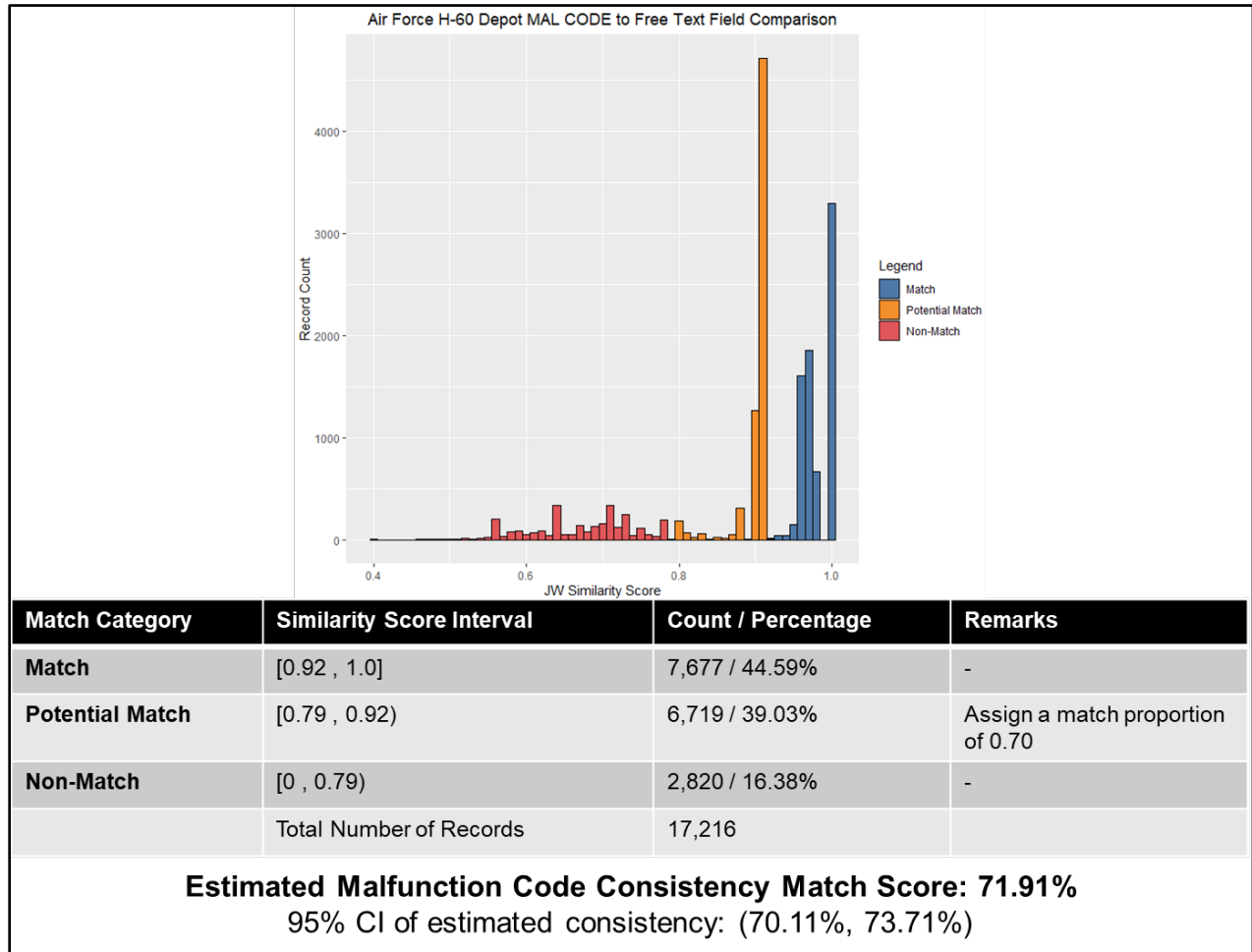


Figure 28: Air Force H-60 Depot Maintenance Data Malfunction Code to Free Text Comparison

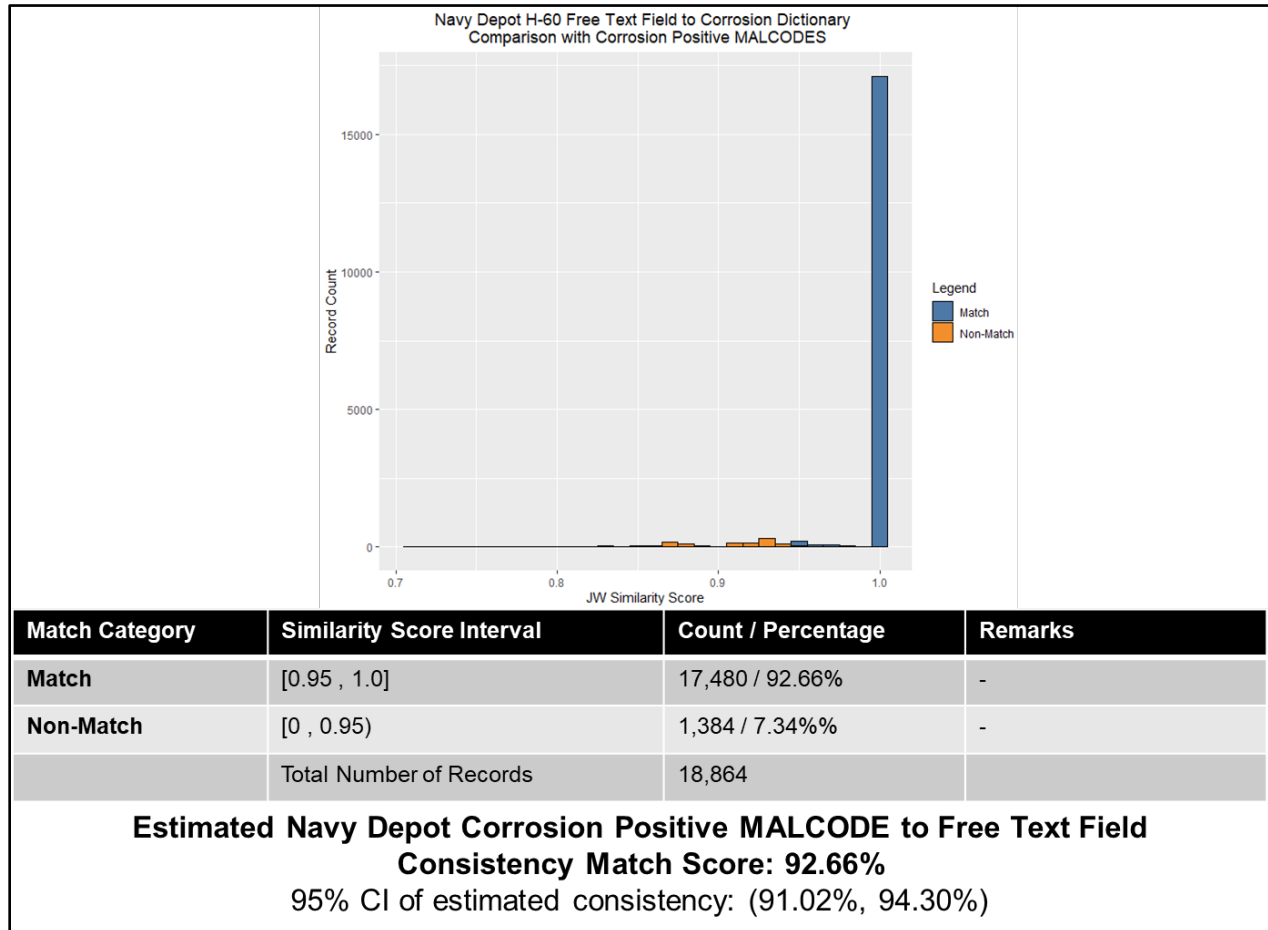


Figure 29: Navy H-60 Depot Corrosion-Positive Malfunction Code to Free Text Comparison

Navy Depot Corrosion MALCODE to Free Text Field Consistency Error Matrix		
		Free Text Field Compared to Corrosion Keywords
		Positive Corrosion Free Text Field Negative Corrosion Free Text Field
CODED FIELDS	Positive Corrosion MALCODE	17,480 1,384
	Negative Corrosion MALCODE	- -
$accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{17,480}{17,480 + 1,384} = 92.66\%$		
$error\ rate = \frac{FP+FN}{TP+TN+FP+FN} = 1 - accuracy = 1 - 92.66\% = 7.34\%$		
$precision = \frac{TP}{TP+FP} = \frac{17,480}{17,480 + 1,384} = 92.66\%$		
$recall = \frac{TP}{TP+FN} = \frac{17,480}{17,480} = 100.00\%$		
$F_1\ score = \frac{2 * precision * recall}{recall + precision} = \frac{2 * TP}{2 * TP + FP + FN} = \frac{2 * 17,480}{2 * 17,480 + 1,384} = 96.19\%$		

Figure 30: Navy H-60 Depot Corrosion-Consistency Error Matrix

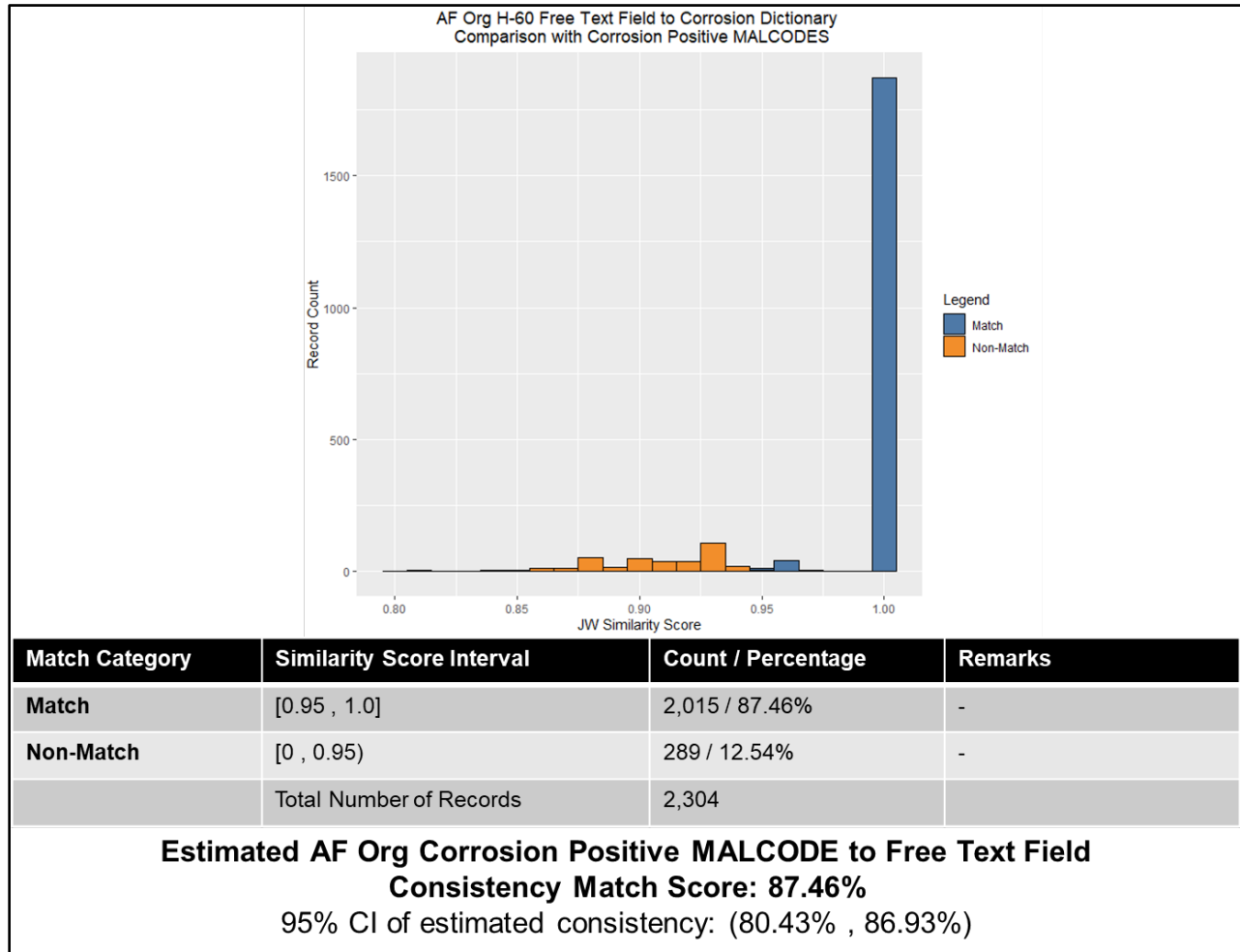


**Figure 31: Air Force H-60 Organizational Corrosion-Positive Malfunction Code to
Free Text Comparison**

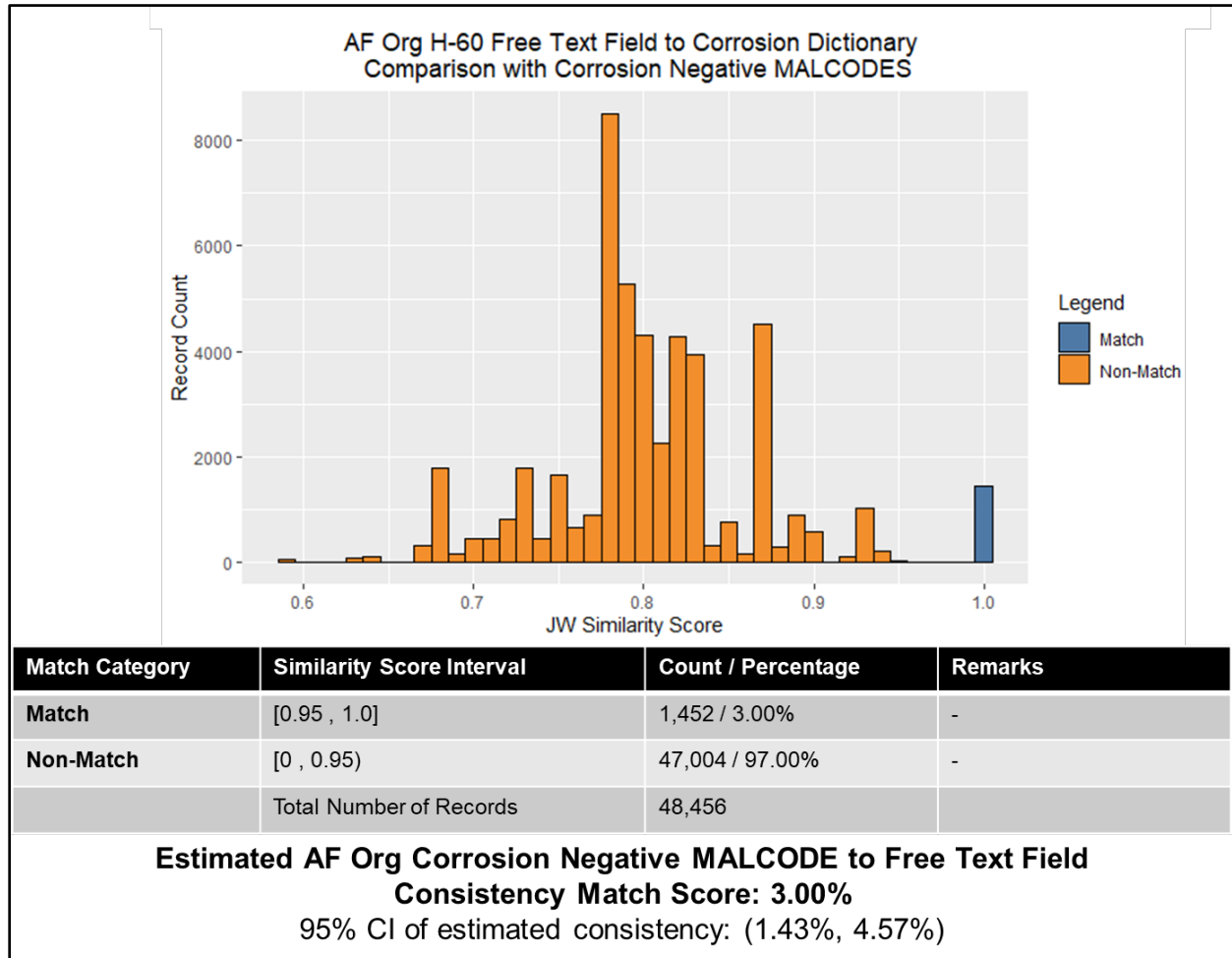


**Figure 32: Air Force H-60 Organizational Corrosion-Negative Malfunction Code to
Free Text Comparison**

Air Force Organizational Corrosion MALCODE to Free Text Field Consistency Error Matrix

		Free Text Field Compared to Corrosion Keywords	
		Positive Corrosion Free Text Field	Negative Corrosion Free Text Field
CODED FIELDS	Positive Corrosion MALCODE	2,015	289
	Negative Corrosion MALCODE	1,452	47,004

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{2,015 + 46,329}{2,015 + 46,329 + 289 + 1,452} = 96.57\%$$

$$error\ rate = \frac{FP+FN}{TP+TN+FP+FN} = 1 - accuracy = 1 - 96.57\% = 3.43\%$$

$$precision = \frac{TP}{TP+FP} = \frac{2,015}{2,015 + 289} = 87.46\%$$

$$recall = \frac{TP}{TP+FN} = \frac{2,015}{2,015 + 1,452} = 58.12\%$$

$$F_1\ score = \frac{2 * precision * recall}{recall + precision} = \frac{2 * TP}{2 * TP + FP + FN} = \frac{2 * 2,015}{2 * 2,015 + 289 + 1,452} = 69.83\%$$

Figure 33: Air Force H-60 Organizational Corrosion-Consistency Error Matrix

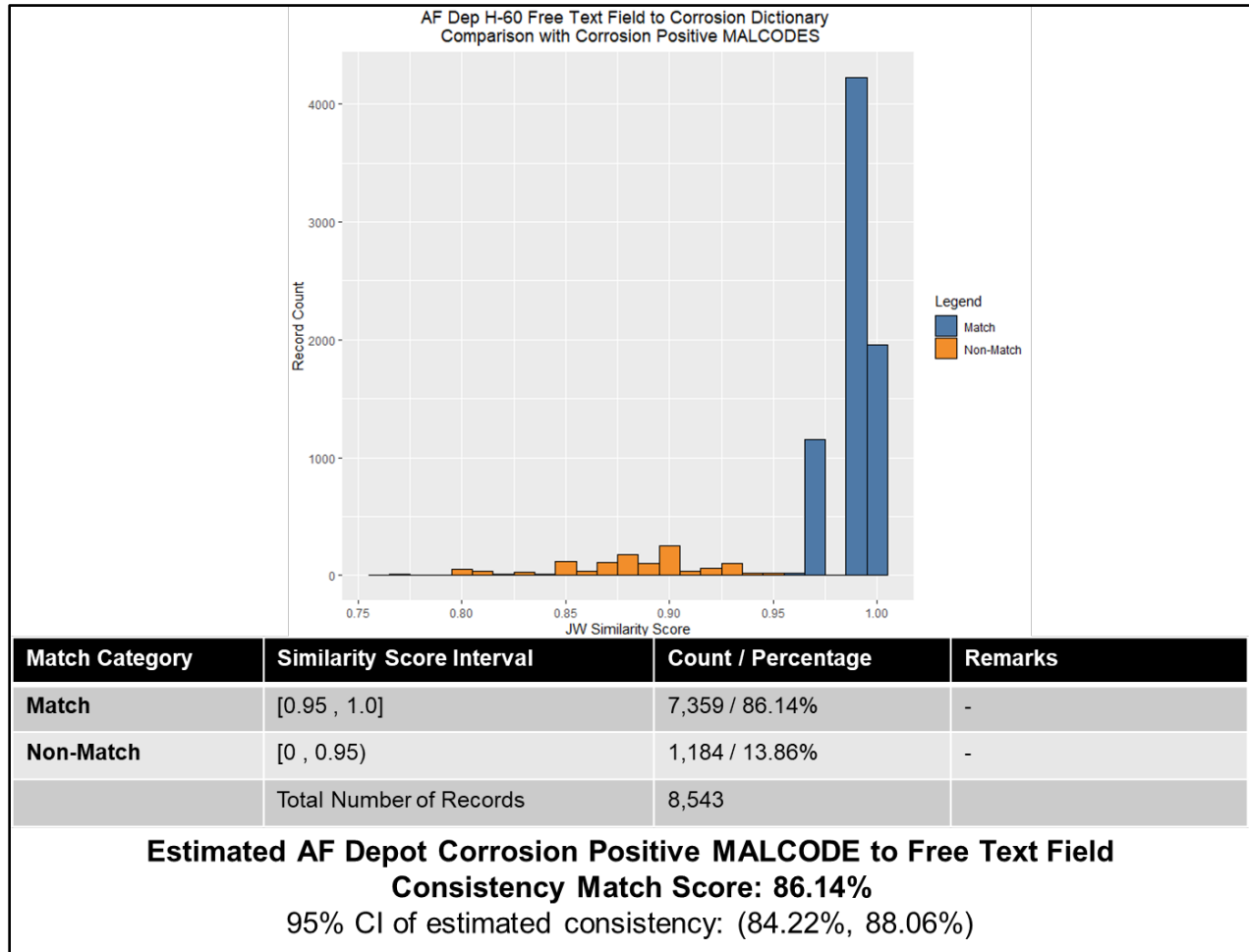


Figure 34: Air Force H-60 Depot Corrosion-Positive Malfunction Code to Free Text Comparison

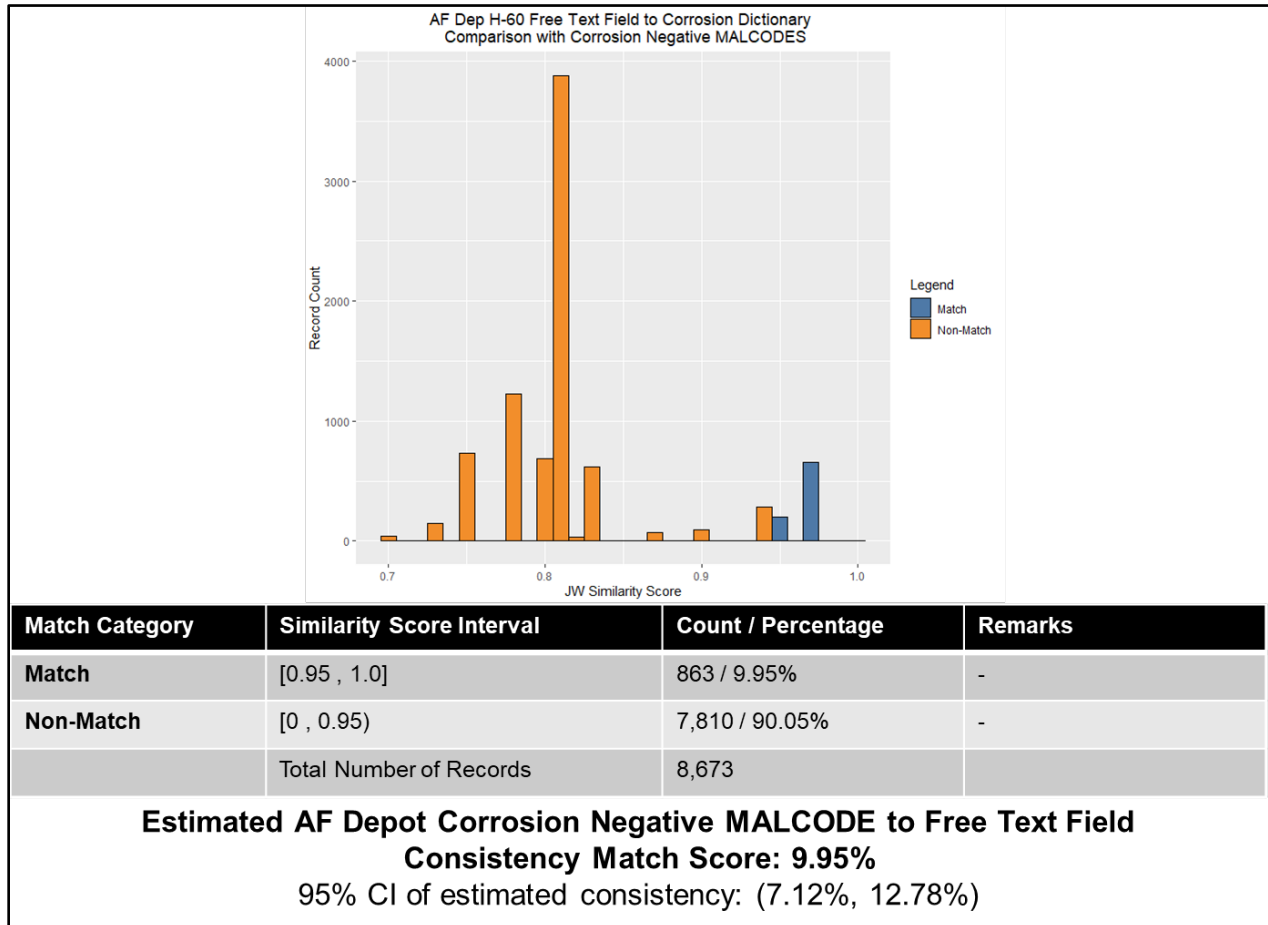


Figure 35: Air Force H-60 Depot Corrosion-Negative Malfunction Code to Free Text Comparison

Air Force Depot Corrosion MALCODE to Free Text Field Consistency Error Matrix

		Free Text Field Compared to Corrosion Keywords	
		Positive Corrosion Free Text Field	Negative Corrosion Free Text Field
CODED FIELDS	Positive Corrosion MALCODE	7,359	1,184
	Negative Corrosion MALCODE	863	7,810

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{7,359 + 7,810}{7,359 + 7,810 + 1,184 + 863} = 88.11\%$$

$$error\ rate = \frac{FP+FN}{TP+TN+FP+FN} = 1 - accuracy = 1 - 88.11\% = 11.89\%$$

$$precision = \frac{TP}{TP+FP} = \frac{7,359}{7,359 + 1,184} = 86.14\%$$

$$recall = \frac{TP}{TP+FN} = \frac{7,359}{7,359 + 863} = 89.50\%$$

$$F_1\ score = \frac{2 * precision * recall}{recall + precision} = \frac{2 * TP}{2 * TP + FP + FN} = \frac{2 * 7,359}{2 * 7,359 + 1,184 + 863} = 87.79\%$$

Figure 36: Air Force H-60 Depot Corrosion-Consistency Error Matrix

Appendix G ADDR Process Flow Chart

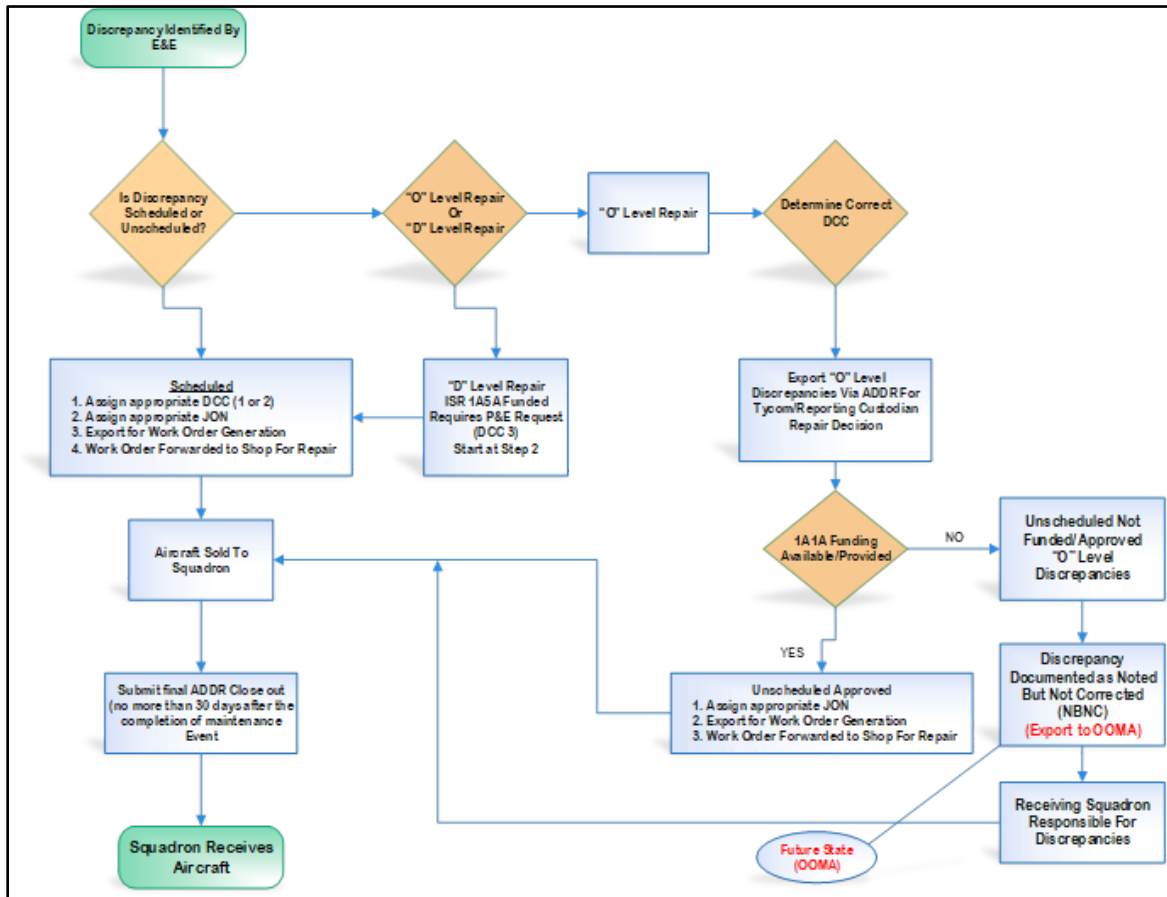


Figure 37: ADDR Process Flow Chart

Appendix H

HMMWV Coded Field to Free Text Comparison

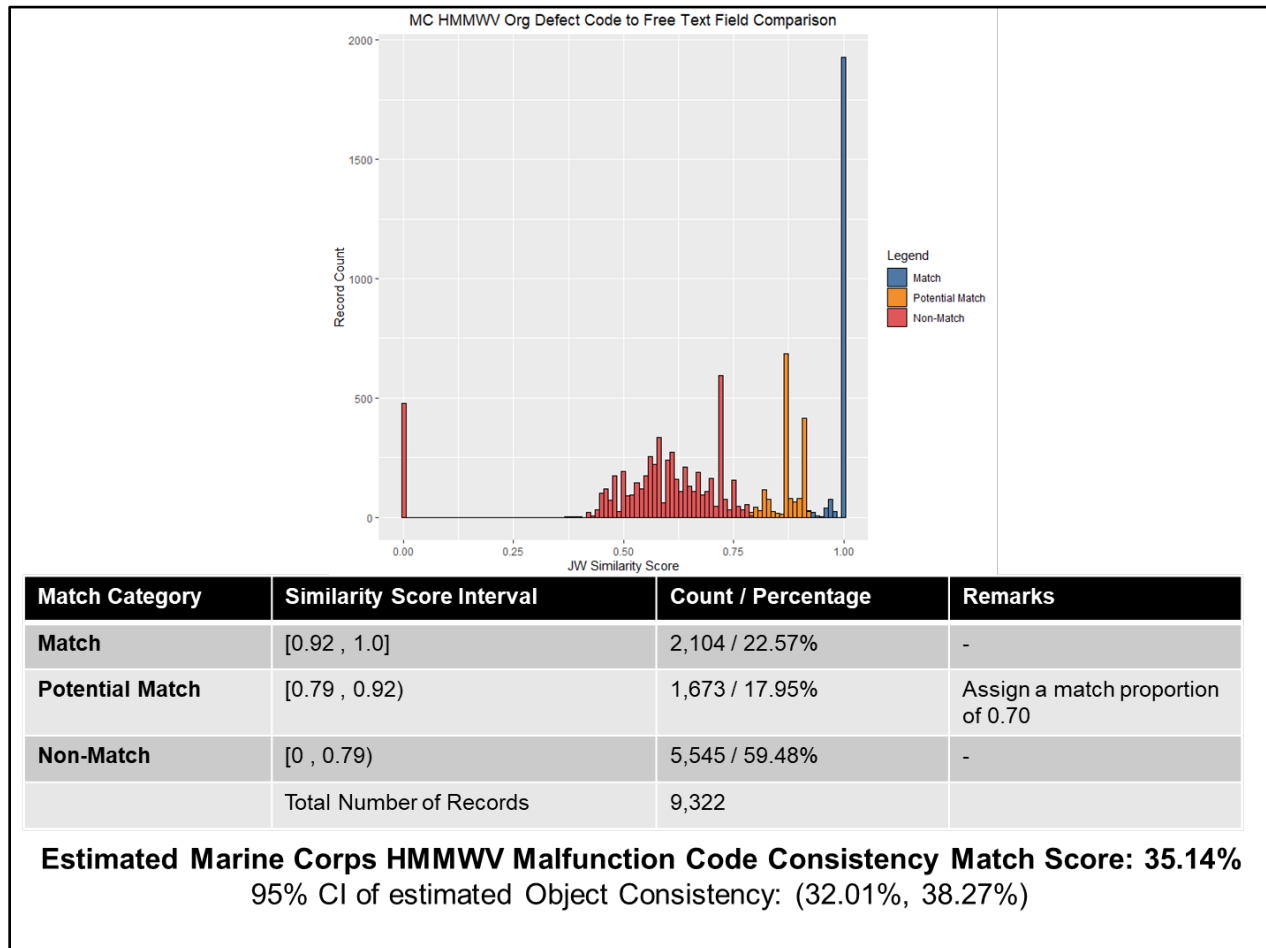


Figure 38: Marine Corps HMMWV Organizational Maintenance Data Malfunction Code to Free Text Comparison

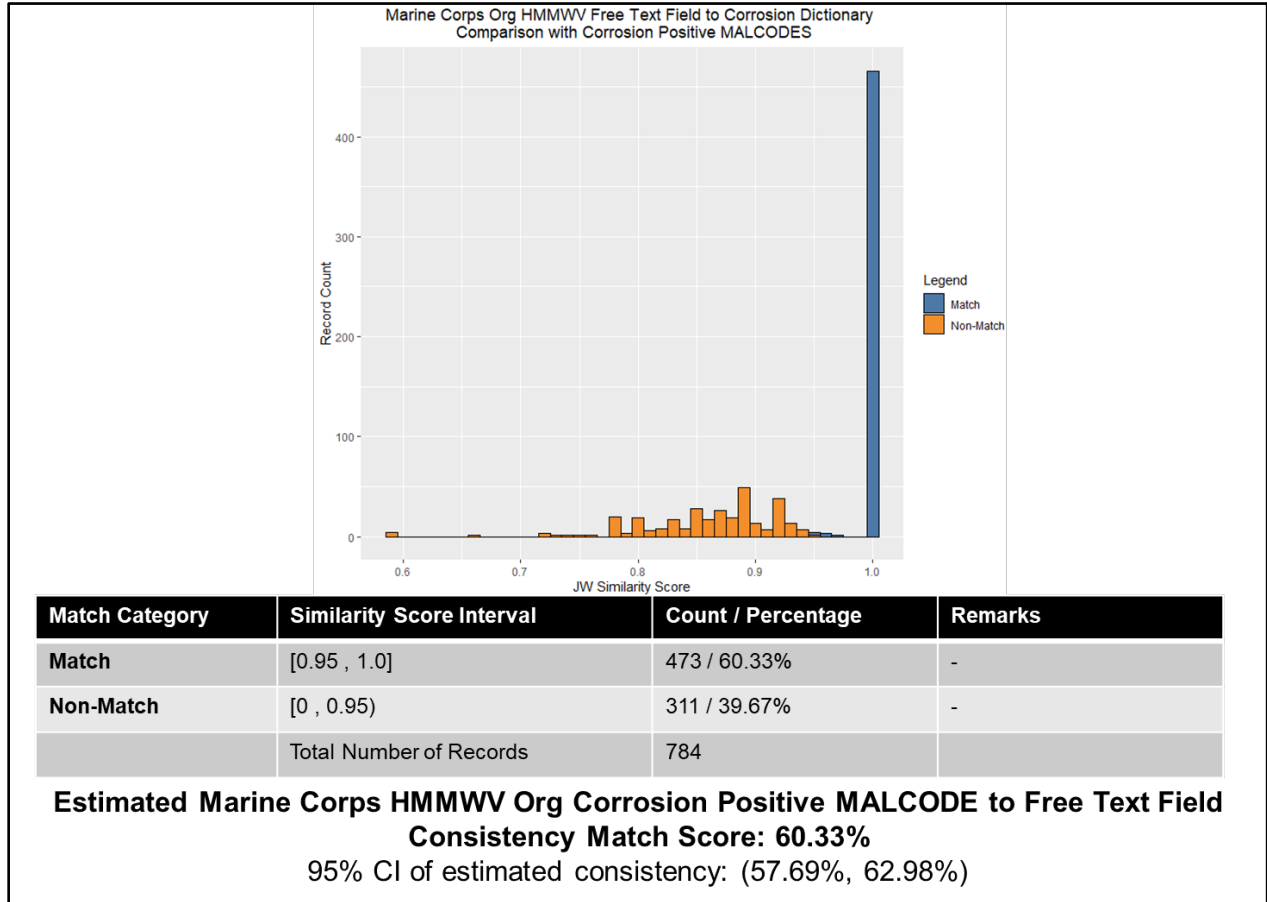


**Figure 39: Marine Corps Organizational Corrosion-Positive Malfunction Code
to Free Text Comparison**

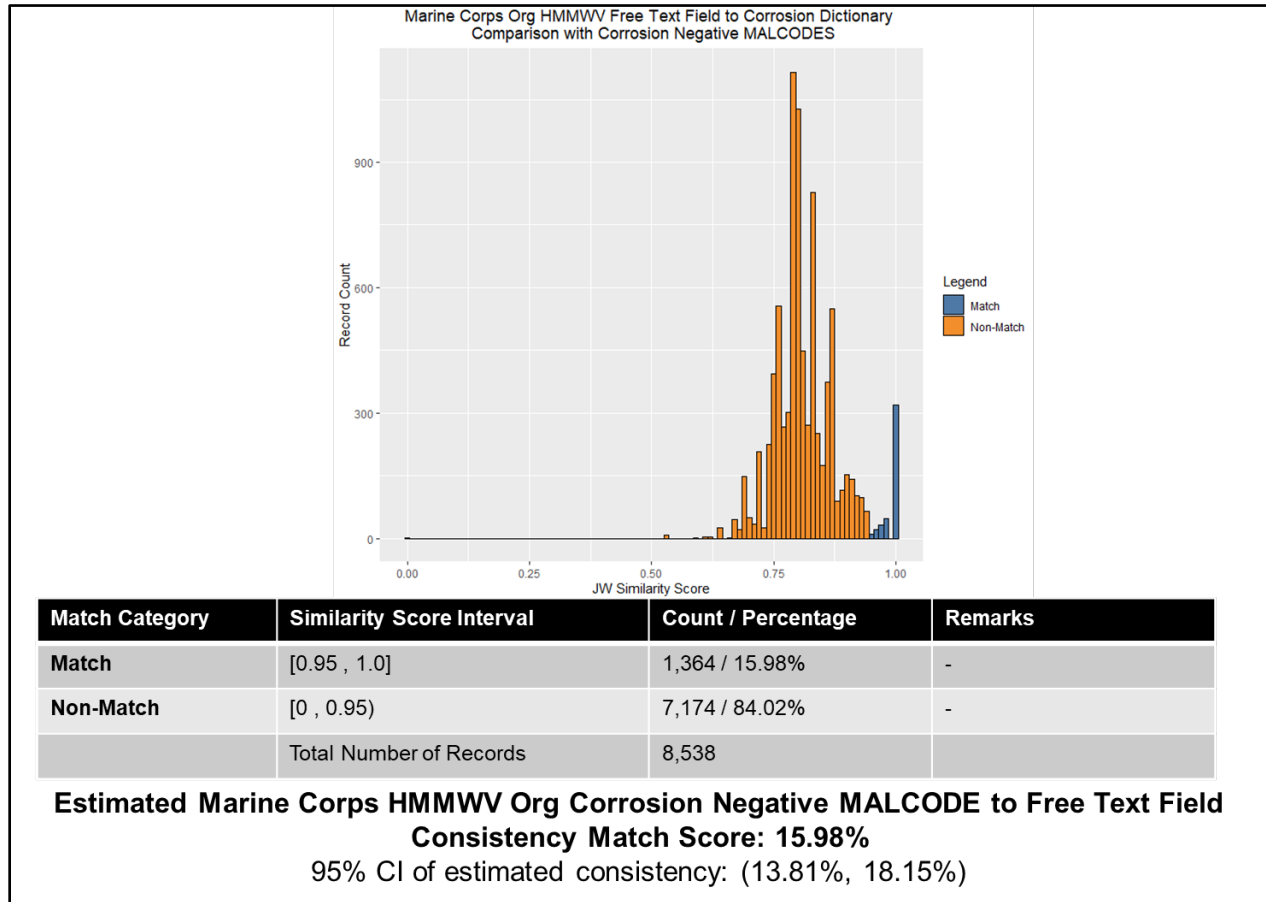


Figure 40: Marine Corps Organizational Corrosion-Negative Malfunction Code to Free Text Comparison

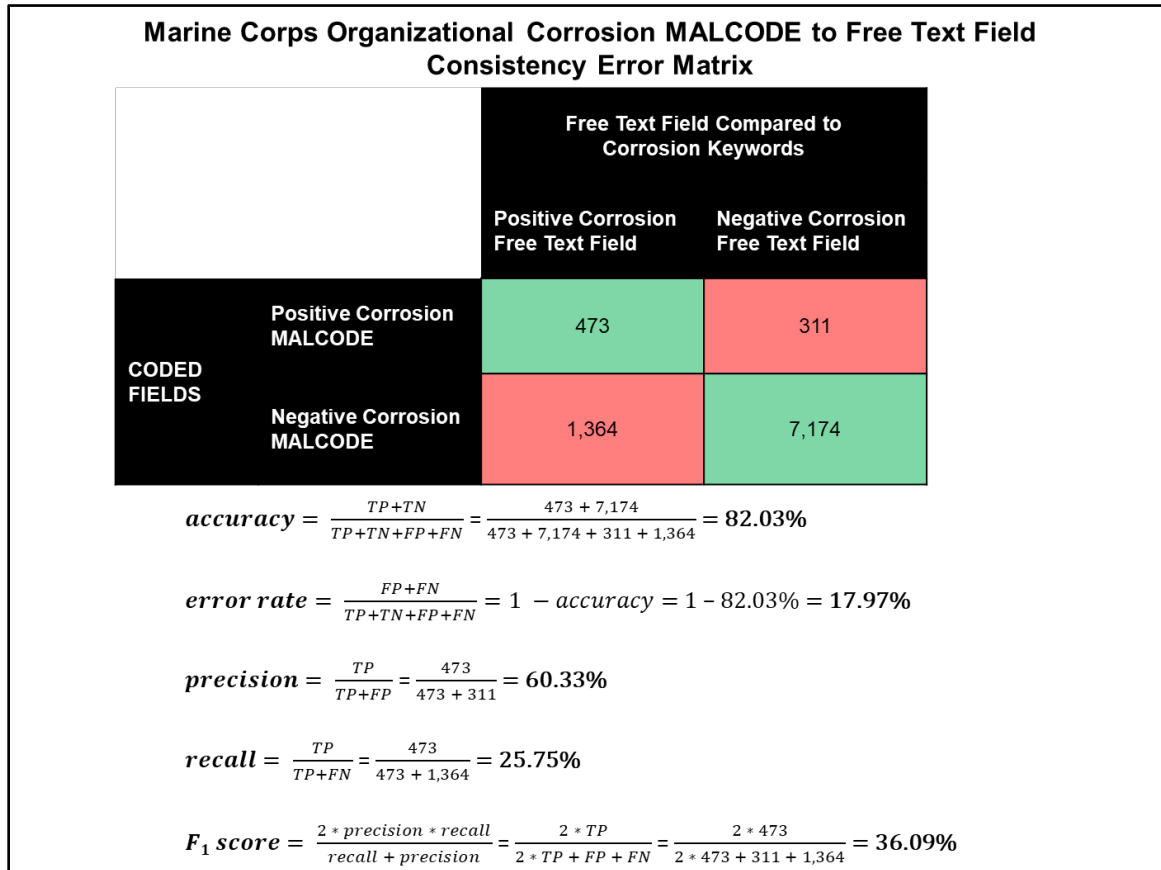


Figure 41: Marine Corps HMMWV Organizational Corrosion-Consistency Error Matrix



Department of Systems Engineering
United States Military Academy
West Point, New York 10996