WHITE PAPER

SUPPORTING VHA IDENTITY MANAGEMENT THROUGH AI

SUBMITTED BY:

Benjamin R. Ambrose, MBA, PCM, CSM,
Program Manager,
Aptive Resources

aptive

TABLE OF CONTENTS

EHRM HC IdM Support Solutioning

Introduction

The Healthcare Identity Management (HC IdM) program, under the Veterans Health Administration (VHA), is responsible for managing identity data in Veteran records to ensure accuracy and integrity across VA systems nationwide. HC IdM tracks a Veteran's service, care, and benefits using a unique identifier known as the Integration Control Number (ICN). This identity data is stored in the Master Person Index (MPI), a repository containing over 75 million unique entries from VA facilities. Together, the ICN and MPI enable HC IdM to manage identity data effectively.

However, the analysis work often involves manually correcting records, which can be time-consuming, especially when resolving complex cases. Complex records may include identity data intermingled across two or more person records, requiring detailed investigation to resolve. Machine Learning (ML) has the potential to streamline this process by reducing the manual workload and accelerating case resolution.

Adding to the complexity, discrepancies in identity data between VA and Department of Defense (DoD) systems present further challenges. This issue is particularly critical as VHA transitions its healthcare sites to the new Electronic Health Record Modernization (EHRM) system. Timely resolution of these anomalies is essential for the smooth operation of healthcare systems and the success of data migration efforts.

Leveraging automation offers an innovative and transformative approach to resolving complex data anomalies. If properly employed, ML can expedite the data correction processes, enhance accuracy, and allow caseworkers to focus on more complex anomalies at an exponentially increased rate. By collaborating with experts in the field of automation and ML the existing HC IdM team aims to significantly improve the efficiency, accuracy, and precision of identity data management– addressing both the current backlog and the growing demands of the EHRM transition. Although significant progress has been made in reducing the number of identity issues shared between VA and DoD, there are a tremendous number of anomalies that remain. Successfully resolving anomalies in discrepant person records according to the current schedule will likely require manual analysis and corrections beyond the ability of the current HC IdM team.

Problem Statement:

The most significant challenge HC IdM faces is the timely resolution of data anomalies, a problem that ML is uniquely positioned to address. The data within the HC IdM ToolKit is sourced from a wide range of contributors including internal health care sites, official government and military documents, and information provided directly by Veterans. The manual processing of person records is complicated by inconsistent data sources that make the timing of resolving inconsistencies even more important. These sources often fail to follow best practices or provide accurate information about person traits, which increases the delays in data resolution. If the HC IdM personnel who are currently focused on anomaly cleansing were to have increased automated processes, then resources could be redirected to resolve the more complex anomalies and redeployed to complete more difficult data cleansing tasks. These adjustments would effectively alter the prioritization of the anomalies in the system thus increasing HC IdM's capabilities.

While matching algorithms address some anomalies, many require manual intervention. This manual resolution can take 10 to 40 minutes to resolve, based on findings from tracking the work speeds of the current project team in training. Data discrepancies are worked, in general, in the order in which they are received. There is an option for staff at health care centers to mark a discrepancy as 'high priority', meaning the discrepancy is affecting a patient currently receiving care or who has an appointment within the next 24 hours. However, sometimes discrepancies are marked as 'high priority' when they do not fall within the criteria. Mixed practices and unclear information from leadership at VHA and community facilities make the use of the 'high priority queue' moot. This mixed prioritization leads to delayed processing of critical cases due to various factors such as caseworker preference, unclear prioritization practices, and siloed leadership guidance. This results in a backlog of critical cases.

HC IdM's data anomalies can be divided into simple and complex cases. Robotic Process Automation (RPA) is ideal for handling simple cases, freeing up caseworkers to focus on more complex issues. Simple cases involve tasks like adding DEERS or SSA correlations to records after verifying the match. RPA is ideal for handling simple cases as it will free up caseworkers to work on more complex issues.

In contrast, complex cases with data overlays involve intermingled identity data, such as a dependent's information appearing in the Veteran guardian's record. These cases require manual investigation and often coordination with DoD, which can extend resolution from days to weeks. By using ML and automation to resolve simple anomalies, HC IdM can reduce backlogs and allow caseworkers to focus on resolving the more intricate, time-consuming cases.

Ultimately, the time spent on manual resolutions directly impacts the efficiency of the system—and by extension, the care Veterans receive. Reducing the time to resolve these anomalies is not just a matter of operational efficiency; it ensures Veterans can receive timely and accurate health care, presenting potential delays in care delivery. Delays in processing records lead to postponed health care and benefits for Veterans and beneficiaries.

Proposed Solution:

Currently, Team Aptive addresses anomalies in data elements between VHA and DoD records created by manually researching identity data sources to discern what has happened to cause the discrepancy. This is followed by integrating the data with DoD into a single Federal Electronic Health Record (FEHR). Rectifying the errors within the records transfers requires extensive algorithmic audit procedures and requires individual caseworkers to investigate current system errors, track data message redirection

and corrections, and contribute to workflow redesign. HC IdM needs a solution that integrates ML and automation into the existing FEHR- one that seamlessly integrates into HC IdM's existing workflows and focuses on compatibility, scalability, and ease of implementation to create seamless continuity of remediation of identification data anomalies.

Team Aptive proposes a phased approach for an efficient and comprehensive solution, with a heavy focus primarily on automation and adding exponential impact to Veterans' lives.

Phase 1

Establishing a Robust Data Foundation for Automation

- Data Collection and Integration
- Anomaly Detection and Definitions
- Data Cleansing and Transformation
- Preliminary Analysis and Insights

Phase 2

Enhanced Automation and Process Optimization

- Implementation of RPA
- Integration of Machine Learning (ML)
- Automating Data Transformation and Validation
- Continuous Monitoring and Feedback

Phase 3

Full Automation Integration into Operational Workflows

- Embedding ML Models into Live Systems
- Continuous Refinement of Models
- Alignment with Operational Workflows
- Utilization of Deep Analytics
- Analytical Dashboards for Real-Time Insights

Phase 1: Establishing a Robust Data Foundation for Automation Implementation

In this initial phase, Team Aptive will focus on building a strong, comprehensive data foundation, essential for the successful deployment of RPA and ML, or other forms of automation. Our goal is to systematically define, gather, organize, and analyze data within HC IdM systems, identifying critical gaps, inconsistencies, and opportunities for improvement that could hinder future automation efforts. By prioritizing data integrity upfront, we set the stage for seamless automation that will speed up the processing of records making it possible for VA to more rapidly deliver benefits to Veterans without unnecessary delays.

Key activities during this phase include:

1. **Data Collection and Integration**

   Team Aptive will extract and integrate data from multiple sources within HC IdM systems including health care records, personal identifiers such as Social Security Numbers (SSNs), EDIPIs, ICNs, and vital statistics such as birth and death dates. By pulling together a comprehensive and holistic dataset, the full spectrum of patient information is available for analysis, minimizing the risk of data fragmentation or omission. This unified approach forms the basis for accurate anomaly detection and prepares the system for future automation processes.

2. **Anomaly Detection and Definitions**

   Using pattern recognition techniques, we will identify and flag anomalies within the dataset such as duplicate entries, conflicting patient identifiers, or data entry errors. A key component of this phase will

be defining the difference between simple and complex anomalies. While we qualitatively understand the nature of these anomalies, our goal is to use all available data to create a quantitative threshold for categorizing anomalies thus allowing for data-driven decisions to be made that maximizes the allocation of resources for resolution.

Bottleneck Identification: Currently, the limited number of government-trained personnel tasked with addressing complex anomalies creates a significant bottleneck. These specialists are responsible for resolving the most intricate issues, but the workforce is too small to meet demand efficiently. By better defining simple versus complex anomalies through data analysis, we can improve resource allocation and streamline the resolution process, therefore reducing delays and increasing overall efficiency.

3. **Data Cleansing and Transformation**

   Once anomalies are detected, Team Aptive will implement data cleansing processes to correct inaccuracies. This will involve standardizing formats (e.g., aligning all date entries to a consistent format), resolving discrepancies in patient identifiers, and ensuring data from different sources is aligned and consolidated. Duplicate entries will be merged, missing fields will be populated through cross-referencing external data sources, and incomplete records will be filled in using reliable data. These efforts will enhance data quality and prevent recurring anomalies, laying a solid foundation for future automation.

4. **Preliminary Analysis and Insights**

   As data is cleansed and anomalies are resolved, we will perform preliminary analysis to extract insights and identify trends. This phase will help uncover patterns in the data, highlighting immediate issues and long-term areas for improvement. These insights will directly inform our approach to RPA in Phase 2, ensuring that automation is tailored to the specific needs of the HC IdM data environment. By establishing benchmarks for anomaly resolution rates, we will gain a clearer understanding of resource needs moving forward.

Immediate Impact to the VHA HC IdM team in Phase I:

- **Improved Data Integrity**

  By cleansing and organizing data from the outset the HC IdM system becomes more reliable, leading to fewer disruptions in identity management processes. This ensures that the VA can provide more timely and accurate health care services to Veterans, even before automation is fully implemented.

- **Reduced Bottlenecks**

  By quantifying the complexity of anomalies and improving how simple versus complex cases are defined, the VA will gain a clearer picture of its resource allocation needs. This helps alleviate the bottleneck caused by the limited number of government-trained personnel working on complex anomalies. Through better data priority processes via understanding the simplicity vs complexity and the minute details of existing and future data anomaly collection, the VA can focus resources more efficiently, speeding up anomaly resolution and improving operational flow.

- **Actionable Insights for Future Phases**

  The preliminary analysis performed in Phase 1 offers immediate insights into data trends and anomaly patterns allowing the VA to make informed decisions about how to approach RPA in the next phase. This lays the groundwork for more targeted automation efforts increasing the likelihood of a successful RPA and ML implementation.

- **Foundation for Improved Veteran Services**

  The immediate improvements in data quality will have a direct impact on Veterans' access to health care services, reducing errors and delays caused by inaccurate or fragmented identity data. Veterans

will benefit from more accurate records, quicker resolution of identity-related issues, and faster access to care.

Phase 1 is designed to deliver immediate and tangible benefits to the VA while setting the stage for more advanced automation in subsequent phases. By focusing on data integrity, bottleneck reduction, and the clear definition of simple versus complex anomalies, we make sure that the foundation is solid for the automation processes to follow. The insights gained from this phase will allow the VA to make more informed decisions about resource allocation and automation implementation, improving overall system performance and helping to safeguard Veterans' access to essential health care services.

Phase 2: Enhanced Automation and Process Optimization

In this phase, the team shifts focus to accelerating the resolution of data anomalies through the strategic implementation of RPA and ML. This phase is designed to enhance efficiency by automating routine data-cleansing tasks, thereby significantly reducing the manual workload for caseworkers and enabling them to concentrate on more complex issues.

Key components of this phase include:

1. **Implementation of RPA**

   The deployment of RPA will automate repetitive tasks such as identifying, flagging, and categorizing anomalies in the data. By streamlining these routine operations, RPA will reduce the risk of human error, standardize workflows, and dramatically increase the speed at which data anomalies are addressed. This automation is particularly effective in handling large volumes of data, allowing for quicker turnaround times without compromising accuracy. For example, RPA will automatically detect duplicate entries, misaligned identifiers, or missing data, allowing caseworkers to focus on more nuanced issues that require expert analysis.

2. **Integration of Machine Learning**

   To enhance the capabilities of RPA, we will integrate ML algorithms to improve the accuracy and intelligence of anomaly detection. These ML algorithms will analyze historical data and identify patterns to refine the detection process to reduce false positives and make the system more effective over time. The ability of ML to "learn" from past data will allow for smarter identification of data anomalies, ensuring that the system becomes increasingly adept at recognizing and resolving issues. This synergy between RPA and ML will create a comprehensive, multi-layered approach to managing data inconsistencies, improving both speed and precision.

3. **Automating Data Transformation and Validation**

   In addition to detecting and categorizing anomalies, RPA will also automate key steps in data transformation and validation. This process includes standardizing data formats, resolving inconsistencies between data fields, and validating entries against predefined rules and criteria. For example, RPA will ensure that all dates follow a standardized format and that personal identifiers, such as Social Security Numbers (SSNs), are validated for accuracy. This automation not only accelerates data preparation but also ensures that the data is reliable and ready for further analysis or integration into other systems.

4. **Continuous Monitoring and Feedback**

   As RPA and ML work together to process data anomalies, Team Aptive will establish a continuous monitoring system to track performance and outcomes in real time. This feedback loop will allow us to refine the automation processes as data patterns evolve. For instance, if certain types of anomalies become more prevalent, we can adjust the system's parameters to address them more effectively. Real-

time monitoring will also enable the team to identify areas for improvement quickly to ensure that the system remains flexible, responsive, and aligned with the HC IdM's operational needs.

By integrating RPA with ML, this phase aims to process vast volumes of data more rapidly and accurately, empowering the HC IdM system to resolve data anomalies with greater efficiency. This not only enhances the integrity of the data but also ensures that Veterans receive timely access to health care services, addressing critical issues that may have previously hindered their care.

Phase 3: Full Automation Integration into Operational Workflows

In the final phase, Team Aptive focuses on seamlessly integrating ML solutions into HC IdM's existing operational workflows. This integration ensures that the resolution of data anomalies becomes an intrinsic part of daily operations, fostering an environment that is efficient, scalable, and sustainable.

Key objectives for this phase include:

1. **Embedding ML Models into Live Systems**

   In this phase, Team Aptive will implement fully developed ML models directly into HC IdM's operational frameworks. This real-time integration will enable the immediate detection and resolution of data anomalies as they arise, vastly improving responsiveness and service delivery. ML will not only allow for proactive issue identification but also automate the resolution process so that data remains accurate and reliable across all systems. This real-time capability means anomalies that previously delayed services will now be addressed swiftly, preventing disruptions and enhancing the overall experience for Veterans.

2. **Continuous Refinement of Models**

   As HC IdM continues to collect new data and generate performance metrics, Team Aptive will continuously refine and improve the ML models. This adaptive approach ensures that the ML solutions evolve alongside the data patterns and operational needs of HC IdM. By regularly analyzing feedback and making necessary adjustments, we guarantee that the models remain effective and responsive to changes, further increasing the accuracy and efficiency of anomaly resolution. This iterative process allows the system to improve over time, offering long-term benefits to HC IdM's operations.

3. **Alignment with Operational Workflows**

   A crucial component of Phase 3 is ensuring that the new ML-driven processes are seamlessly aligned with existing HC IdM workflows. This alignment fosters collaboration across teams and departments, helping to ensure that the integration of ML enhances—not disrupts—ongoing operations. Our focus will be on embedding ML tools in a way that complements current practices while improving the speed and accuracy of data management. This ensures that the benefits of ML are fully realized without creating additional complexity in the day-to-day operations of HC IdM personnel.

4. **Utilization of Deep Analytics**

   One of the most powerful capabilities of ML is its use of deep analytics to uncover underlying trends and potential gaps in data. By harnessing this capability, Team Aptive will enable HC IdM to proactively correct identity data errors before they escalate into larger issues. The predictive power of ML will allow for better preparation for "go-live" events and other critical moments, ensuring that HC IdM is operationally ready at all times. This proactive approach will greatly reduce backlogs and improve the overall accuracy of data management.

5. Analytical Dashboards for Real-Time Insights

To further enhance decision-making, Team Aptive will develop analytical dashboards that visualize data insights derived from ML-driven pattern recognition. These dashboards will provide real-time updates on completed and outstanding data anomalies, giving HC IdM leadership and caseworkers the ability to monitor progress and performance at a glance. The dashboards will also help in prioritizing cases so that the most critical issues are addressed promptly. This visibility will foster transparency and enable faster, data-driven decision-making throughout the organization.

By embedding ML models into live systems, HC IdM will be able to detect and resolve anomalies as part of its routine operations, resulting in a dramatic reduction in backlog and improving service quality. For example, anomalies related to duplicate patient records or incorrect personal identifiers can now be flagged and corrected automatically, reducing the time needed to resolve such issues manually. Furthermore, by aligning these processes with operational workflows, HC IdM staff will benefit from a streamlined system that reduces errors and enhances overall productivity.

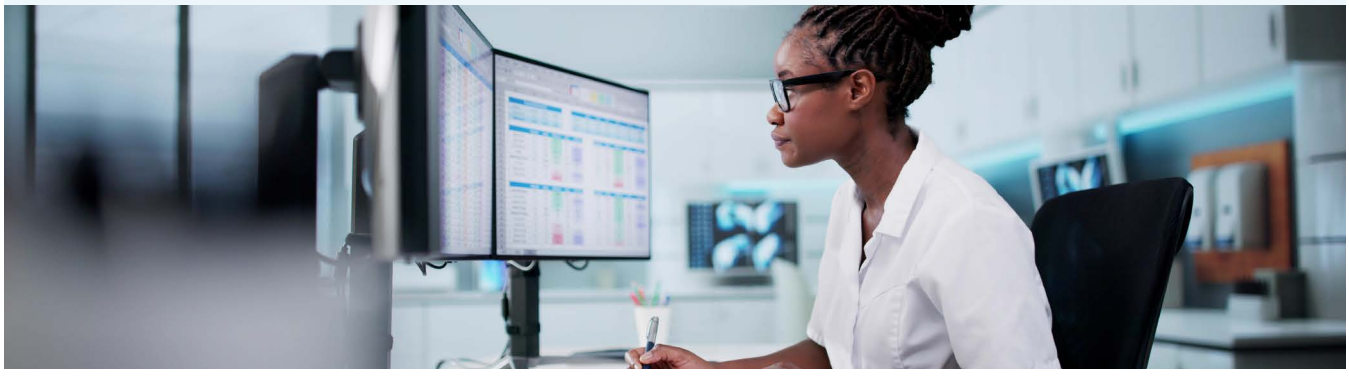
Benefits/Expected Outcomes:

By increasing the use of algorithmic solutions for HC IdM processes, we can immediately accelerate the speed at which data anomalies are resolved. Currently, it takes trained contractors an average of 41.5 minutes to resolve a single anomaly. Implementing this methodology will streamline that process, ensuring that Veterans and their beneficiaries can receive critical services—such as medical treatment, financial benefits, and other essential support from the VA—without unnecessary delays.

In addition to the immediate improvements in efficiency, this approach will maintain a 100% accuracy rate, which means that every anomaly resolved will lead to more reliable and trustworthy data across the system. Veterans who have experienced delays in care or benefits due to data discrepancies will see a marked improvement in the speed at which these issues are corrected, leading to tangible and positive quality-of-life changes.

Hundreds of thousands of Veterans have had their medical care or benefits impacted by data errors. Addressing these discrepancies with increased automation not only ensures timely access to care but also helps fulfill the government's responsibility to deliver the highest standard of service to Veterans. By tackling these issues now, we lay a solid foundation for the future automation of simple anomalies allowing more complex cases to receive the focused attention of highly trained personnel.

Moreover, resolving anomalies more efficiently improves the real-time exchange of data between systems, makes data flows more reliable, and enhances decision-making processes across the VA. By adopting this methodology, the VA remains at the forefront of medical care and technological innovation and solidifies its role as a leader in health care innovation and artificial intelligence (AI) advancement. This approach not only benefits Veterans but also paves the way for future enhancements in hospital systems nationwide and reinforces the VA's commitment to innovation and cutting-edge solutions.



Conclusion: Advancing Veteran Care through Enhanced Identity Management

Veterans who have dedicated their lives in service to our nation deserve timely and accurate access to the health care and benefits they have earned. However, due to identity management data discrepancies many Veterans experience delays or complications in receiving these essential services. The current system's limitations create a backlog of unresolved cases that impacts both the efficiency of health care delivery and, ultimately, the well-being of those who rely on the Veterans Health Administration (VHA).

HC IdM plays a vital role in ensuring Veterans and their families receive the care and benefits they are entitled to. By improving the processes that manage identity data we can enhance both the speed and accuracy with which services are delivered. The integration of ML offers a powerful and scalable solution to the challenges facing the HC IdM program and therefore allows us to automate the detection and correction of data anomalies while reducing the manual workload for caseworkers.

Under the guidelines of Executive Order 13960, which promotes the use of trustworthy AI in federal systems, the VHA has taken steps to explore and implement solutions that adhere to the highest ethical standards. Through the application of machine learning, the Office of Healthcare Informatics (OHI) can improve data accuracy and streamline operations, ensuring identity management services are as efficient and effective as possible. ML solutions will enable real-time detection of data issues, continuous improvement of workflows, and better prioritization of critical cases.

By integrating ML-driven analytics into HC IdM operations VHA leadership will gain access to previously unavailable insights. Advanced data visualization and pattern recognition will provide real-time updates improving decision-making and facilitating faster responses to identity data anomalies. This will not only enhance service delivery but will also strengthen the VA's position as a leader in health care innovation.

Implementing ML is not just about resolving current data challenges—it is about building a foundation for the future. With automated processes, HC IdM can continuously improve, ensuring Veterans receive the care and benefits they need without unnecessary delays. As the VHA continues to transition to new health care systems these solutions will provide critical support to help ensure every Veteran's information is accurately managed.

In summary, the integration of ML into HC IdM is a crucial step forward in improving the quality and efficiency of care for Veterans. By addressing data anomalies at scale this solution will enhance Veterans Affairs' ability to meet the growing demands of its health care system, ultimately improving the lives of Veterans and their families. Together, we can ensure that the VHA continues to provide the highest standard of care while embracing the future of health care technology.