AN INTRODUCTORY ANALYSIS OF THE AUTOMATED MANIFEST SYSTEM AND THE GLOBAL TRANSPORTATION NETWORK WITHIN THE JOINT TOTAL ASSET VISIBILITY (JTAV) SYSTEM USING THE DEFENSE LOGISTICS INFORMATION SERVICE DATA QUALITY PLAN

(Research in Progress)

Richard A. Hansen
Defense Logistics Information Service
Richard.a.hansen@dla.mil

Lt. Karen D. Cooper
United States Navy
Karen.cooper@dla.mil

Abstract: This paper presents the Defense Logistics Information Service Data Quality plan. The data quality plan is defined and the action plan for one system is included. It discusses two systems involved in the prototype effort of the data quality plan. The data quality errors are briefly discussed.

Key Words: Department of Defense (DoD), Defense Logistics Information Service (DLA), Defense Logistics Information Service (DLIS), Automated Manifest System (AMS), Global Transportation Network (GTN), Joint Total Asset Visibility (JTAV), Transportation Container Number (TCN)

INTRODUCTION

The Defense Logistics Information Service (DLIS) is a field activity of the Defense Logistics Agency (DLA), a Department of Defense (DOD) Agency whose mission is to manage supplies and supply chains in support of the Military Services and other DOD activities. DLIS manages the catalog system for the Armed Forces. All military supplies and equipment, from nuts and bolts to fighter jets and submarines, are listed in this catalog. The catalog plays a vital role in the huge supply chains that sustain our nation's fighting forces. It serves as the common frame of reference that enables buyers in DOD to communicate with the nation's industrial base that produces the supplies and equipment used by the military. Our role is a fairly narrow one. We do not store or issue items. However, our nation's Armed Forces use information from our databases for virtually all logistics business, be it supply, maintenance, or transportation. Our catalog lists nearly seven million items. We provide over 100 different pieces of information on each item. The information is standardized across all Military Services to provide the utmost in interoperability. Through our significant international role, it also provides for interoperability among NATO members and other allies. Some of the information we maintain is used to help protect the environment and we specifically identify items that are environmentally friendly [2].

BACKGROUND

At the 9th ICIQ, then DLIS Commander Col Joseph Cassel gave the keynote address. In his address he made note that reporting dirty data was ok. To help the program managers realize this DLIS has developed a corporate data quality plan. Developing and implementing a corporate data quality plan will support the vision that DLIS is the premier provider of Department of Defense (DOD) supply chain data and logistics information technology solutions. DLIS is the program manager for approximately 120 DLA systems; however DLIS does not own the data. DLIS is not always the authoritative source of the data it creates and maintains. The concept of data ownership, specifically when the ownership is bestowed on a single individual or department, cultivates the wrong mind set. Ownership implies control, and program management authority to change data at will within the DLIS systems is limited. DLIS is, in actuality, a "data steward", (a trustee). DLIS has been assigned responsibility for the quality of the data but at the same time, has not been given full authority to correct and maintain that data as needed. If there are problems with any of the data, customers will respond to DLIS. Therefore, we have a direct interest in assuring that any data displayed on DLIS systems or products is "quality data". When deficiencies exist with the data we receive and display from external systems, we must consult with the data provider and strive to have the customers' requirements met. Additional factors influencing the implementation of recommendations include: (1) the availability of resources needed to accomplish the improvement, (2) the schedule of software releases, and (3) changes to the Automated Information System (AIS) hardware environment. This plan establishes a process for the overall DLIS data quality. It will define our approaches for the organization, process, technology, cost and challenges to be addressed. The plan establishes the basic methodology which will be used to identify, analyze, measure, and grade individual systems and products on the quality of data. As an organization, we must recognize that working towards data quality is a process to facilitate awareness and encourage growth and improvement. *Assessment of DLIS systems and products may result in reporting less than satisfactory data quality; however it is a method to attract attention to the issues and gather support for the necessary corrections*. The Data Quality process is to be the means to an end rather than the end itself. Many scholars have struggled with the definition of *Data Quality*, it's generally agreed it should be based on the customer's needs and should be defined on those basis by the program manager/data steward. It is necessary, however to include a more comprehensive view that includes a process oriented definition. Our definition of Data Quality is: *The degree to which our data correlates with factual information based on accuracy, consistency, currency, and completeness*. Below is a definition and example of the four attributes that will be used to measure DLIS systems and products.

- **Accuracy**: The measure or degree of agreement between a value (or set of values) and reality. The data is correct for what is being represented.
- **Consistency:** The data passes all edits for acceptability i.e. format –length, characteristics, values; i.e. 3 positions = 3 positions; numeric = numeric; value = is in the table.
- **Currency:** The data is up-to-date and the age of the data is appropriate for the task at hand.
- **Completeness:** The measured data that should have values in them, in fact do so. Input would be based on customer/system needs.

The DLIS corporate board has targeted four systems to prototype using the data quality plan. Other systems will be assigned a data steward and continue through the data quality plan process. At this point the data quality process will have progressed past the prototype stage. The systems are:

- Hazardous Material Information Reporting System (HMIRS)
- **Joint Total Asset Visibility (JTAV)/Asset Visibility (AV)**
- Federal Logistics Information System (FLIS)
- Reference Data Environment (RDE)

THE DATA QUALITY PLAN

The DQP process consists of eight sections. This paper will discuss section III, IV, V, and VI. A brief highlight of sections I, II, VII, and VIII follows. Section I is the Executive Summary. This gives a high level description of the reason for the data quality plan. Section II discusses the purpose, goals and objectives of the data quality plan. Section VII contains the attachments and section VIII is the references.

SECTION III

Section III consists of a six-step structured approach (see figure 1) to improve and manage data quality. Prior to beginning the six step review process, the Environment must be established. The Environment provides direction as to who the key players are and what their responsibilities will be. A Data Steward (DS) and a Program Manager (PM) is assigned to system or product slated for a review. Together the DS and PM perform the six step Data Quality Process. The DS has primary responsibility of steps 1, 2, and 5, while the PM has primary responsibility of Steps 3 and 4. Responsibility of Step 6 is shared by both the DS and the PM.

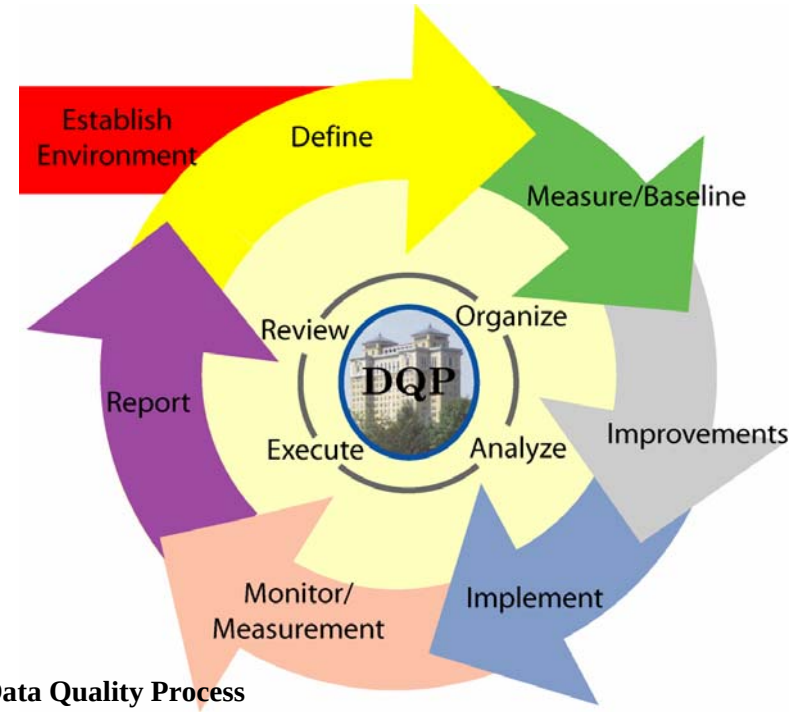

Figure 1 – Data Quality Process

STEP 1: DEFINE

Data quality issues occur in many ways:

- Data entry personnel often fail to follow data entry and maintenance procedures
- Proper system controls may not have been implemented to minimize the possibility of quality problems which involve operational systems that feed into DLIS systems
- Accuracy of business rules and systems edits has not been verified or implemented
- Data being received from outside enterprises may be questionable and extremely difficult to resolve because of little or no control over the source of the data
- Data redundancy introduces the risk of discrepancies when all copies are not kept in sync and current. PM/DSs must be mindful of these issues when defining the scope for systems and products.

With assistance from the PM, the DS begins the define step by documenting the overall scope. System generated exception reports on data that does not conform to established standards and business rules are also included as well as a description of actions needed to improve data quality, the rationale, lessons learned, and improvement metrics. A parallel approach is taken to determine and prioritize the "*Data Quality Issues*" (measurable observations) according to customer impacts, feasibility, and cost-benefit analysis as applicable. Starting with known issues and then moving out with customer surveys.

There is no specific number of *Data Quality Issues* that must be measured. The PM and DS must be able to explain what they are measuring and why. They must also be able to justify why they choose not to measure other potential or common *Data Quality Issues*.

SCOPE OF JTAV

The Joint Total Asset Visibility (JTAV) system takes data from 39 sources and provides a fused view of the data. JTAV provides the Unified Commands the ability to locate assets in storage, in process, in transit, and in theater. JTAV provides users a web access to this information. Additionally, each Command has a JTAV server that supplies data by either reaching back to the source as needed or by storing the data forward in anticipation of users' needs. There are 39 basic data feeds with other multiple data feeds to the basic ones. The data feeds are conducted hourly, daily, weekly, monthly, and quarterly. The data feeds are either automated excel, or flat files which are loaded [4]. The identification of JTAV and the feeder systems is step one of the Data Quality Plan, Define.

A particular problem affecting asset visibility is the incompatibility of container number formats among feeder systems. In particular, Global Transportation Network (GTN) requires a 10-character container number, while the Radio-Frequency In-Transit Visibility System (RFITV) uses 11 characters (the same format as GTN plus a check digit), and AMS requires a 5-character number. Furthermore, some input data includes a dash or space that is read as an extra digit, and other systems drop zeroes located in the last position. Once the data is fed into JTAV, these inconsistencies cause queries to be unable to match these records. This causes a loss of ability to track containers and their contents at a higher level. For this data issue this is step one of the Data Quality Plan, Define.

While GTN is not a DLA managed system the container number issue crosses the bounds between several systems. The U.S. Transportation Command (USTRANSCOM) GTN gives its customers located anywhere in the world a seamless, near-real-time capability to access – and employ – transportation and deployment information. GTN is an automated command and control information system that supports the family of transportation users and providers, both Department of Defense (DoD) and commercial, by providing an integrated system of in-transit visibility information and command and control capabilities. GTN collects and integrates transportation information from selected transportation systems. The resulting information is provided to the SECDEF, Combatant Commanders, USTRANSCOM, its component commands, and other DoD customers to support transportation planning and decision-making during peace and war. In keeping with modern technology, GTN is completely available on the Internet's World Wide Web [3].

SCOPE SUMMARY

The JTAV system receives data feeds from various service systems at various times (quarterly, monthly, weekly, daily, and hourly). The JTAV system has limited data quality checks in place; JTAV is not the owner of the data. The JTAV Action Plan is for the Automated Manifest System (AMS), which is a DLA managed system. It was decided to use a DLA system as the prototype for the JTAV data quality action plan.

DATA SUBJECT AREAS:

º Collectors (Input) These are the input systems to JTAV.

AFEMS-USAF Unit Equipment	**AMS**
ATAV – Army AMMO (Wholesale)	ATAV – Army CL IX Wholesale
ATAV – Army Retail (SARSS)	ATAV – Army Unit Equipment
ATAV AR	ATAV- Army Prepositioned Stocks
ATAV NG	ATAV SO
ATAV / WARS Army Ammo (Retail)	CASEMIS
DAAS – LIPS (LOTS)	DAAS AMMO LIPS
DFAMS	DFW
DISMS	DODAAF
DRMS	DSS
DTRACS	(EUCOM)
DLIS (Formerly FLIS)	FIMARS
GTN	JMAR
MAARS II	MCSCS
RFITV (CONUS)	SAMMS
SASSY - USMC Unit Equipment	SASSY -USMC Retail
SBSS	TAMMIS MEDSUP
U2 (Formerly UADPS)	

º Custodians (Maintenance) JTAV

º Customers (Users) N/A

° Consumers (Products) TBD

DATA QUALITY ISSUES

These are the top five issued identified by the JTAV program management office.

1. Logistics Integrated Data Base (LIDB) (Army Retail): Between October 2003 and January 2004, JTAV experienced an unexplained loss of 75% of Army Retail records.

2. Logistics Integrated Data Base (LIDB) (Army Wholesale): Records are deleted from this dataset only during a monthly reconciliation process; the result is that during the month, JTAV continues to show items that have been deleted in the interim since the last reconciliation.

3. Automated Manifest System (AMS): JTAV has identified a number of inventory distribution sites that are not reported in AMS data; also, there is no reliable unique key to AMS data, which results in commodities in separate shipments appearing to be part of a single shipment.

4. Joint Medical Asset Repository (JMAR) (Wholesale Medical): For more than a year, JMAR has not been able to provide wholesale medical data.

5. Radio Frequency - In-Transit Visibility (RF-ITV) (In-Transit commodities): Inconsistency between the location data of Interrogators in the INTEROGATOR table with the location info located in a free text field in the RF_TAG_TRANSACTION table. When Interrogators are temporarily moved for an Exercise, the location info is updated in the RF_TAG_TRANSACTION table but not in the INTEROGATOR table. However, the text field is not always filled in the RF_TAG_TRANSACTION table.

STEP 2: MEASURE/BASELINE

DLIS has identified four data quality characteristics that will be utilized to measure the quality of our data. The four characteristics are*: Accuracy*, *Consistency*, *Currency*, and *Completeness*. Table 1 illustrates these four Data Quality Characteristics, their definitions, and how each will be measured.

Due to the multivariate nature, it may be necessary to establish weights for each of the four characteristics. This allows the PM and DS to tailor the template to the system/product needs as well as specify the degree of importance each variable has to the overall data quality score or grade. It is possible that one or more of the characteristics are not applicable to a particular system/product. To insure the rating is normalized, each weighting factor should be between zero and one, with the total weight equal to 1.0

Table 1: Data Quality Characteristics

Data Quality Characteristic	Definition	Metric	Weight	Grade (Notional)
Accuracy	The measure or degree of agreement between a value (or set of values) and reality. The data is correct for what is being represented.	Percent of values that are correct when compared to the actual value	Varies according to system Measured as a decimal value	90 – 100% 80 – 89% 70 – 79% 60 – 69% 59% and below
Consistency	Data passes all edits for acceptability. Can often be controlled with some sort of system edits. Data elements are in correct format, length, characteristics, and values.	Percent of data having values that fall within their respective domain of allowable values	Same as above	Same as above
Currency	Up to date	Percent of data that is up to date within a specified threshold time frame	Same as above	Same as above
Completeness	The measured data that should have values in them, in fact, do so. Input would be based on customer/system needs	Percent of data fields that require and have values entered into them	Same as above	Same as above

The percentage values established in the metric will be used to assign a rating for each of the four characteristics. Below is the rating scale and the corresponding colors that will be used (Figure 2).

FIGURE 2

MEASURABLE OBSERVATIONS

Data Quality Issue #3 – **AMS**

The measurable observations come from spreadsheets we receive from an AMS administrator. These measures are using the sum of rows added and updated to a specific table in AMS for a specific period in time. This is a percentage of the sum of the rows listed above versus the total number of rows dropped.

The weighting factors for the AMS Missing Data are: Accuracy .7, Consistency N/A – this attribute was not measured, Currency N/A – this attribute was not measured, and Completeness .3.

The rational for accuracy = .7 is that the information must be accurate for the combatant commanders. The decisions they make are based on where material and personnel are at any given moment. The AMS system is not accurately updated and rows of information are discarded because of missing data then this is an inaccurate process. Based on discussions with an AMS administrator it was decided that accuracy was the most important attribute of the four in this plan.

The rational for not measuring consistency and currency is that the information is feed daily and in a structured format.

The rational for completeness = .3 is that the information needs to be as complete as possible. There are three segment of the data feed and not all segments need to be compete to have a complete data load.

At this time we can not identify which aspects of the system are performing well, and which warrant improvement. There has only been limited information and research conducted of AMS. A statement of work (SOW) has been written and we are waiting for a contractor to assist in the JTAV data quality plan.

When DLIS identified JTAV as a prototype data quality plan system the JTAV PM was contacted. In discussions with the JTAV PM it was decided to concentrate on DLA managed systems (stay in our own backyard). It was decided to use the Automated Manifest System (AMS). AMS is a DLA managed system which provides in-the-box (containers and air pallets) visibility for DLA - Defense Depot Regions

East (DDRE) / West (DDRW) shipments. Data files are pushed from DAASC (AMS sends to DAASC, who then pushes it to the Alexandria production server) [4]. The problem is that some of the records AMS is receiving are not complete. The incomplete records are not loaded/updated in the AMS system, thereby losing visibility in the transportation arena. An AMS system administrator was contacted and the data quality branch started to receive excel files with information relating to the number of missing records. The process identified above is step two of the Data Quality Plan, Measure/Baseline. See chart 1 for information.

Chart 1 AMS Missing Data

JTAV AMS DISCARDS (no GBL_NO)

		approx. # each type record:			comments	ship_from location(s)
		'1'	'2'	'3'		
DATE	TOTAL	AMS_GBL	AMS_SHIPMENT	AMS_LINEITEM		
8/11/2004	131	7	49	75		SW3142,SW3227, SW3216,SW3211
8/12/2004	424	4	280	140		SW3124
8/13/2004	0					
8/14/2004	no info					
8/15/2004	8414	24	1896	6496		SW3224
8/16/2004	0					
8/17/2004	0					
8/18/2004	2					
8/19/2004	3151	52	651	2448		SW3124,SW3117

A JTAV site administrator conducted a limited study of the AMS and highlighted three specific data quality issues:

- Not all DLA facilities use AMS
- The AMS data is at times suspect because of the linking key – GBL_NO – which is not unique
- DLA facilities create RF tags for DVD shipments, but do not create AMS data even though AMS data is created at that facility.

This study assisted the DS in the creation of the JTAV action plan. The action plan is limited in scope because the data is not owned by DLIS or JTAV but by the military services[5]. The JTAV DS is in the process of gathering data to develop a more detailed and achievable action plan using the tools mentioned earlier.

Attachment D Template for Analysis of JTAV Potential Issues (Figure 3) is used to document and display in a quick one look format the necessary information about an issue. The DS along with the PM complete the four sections – DQ Indicator, Root Cause, Current Status, and Proposed Solution.

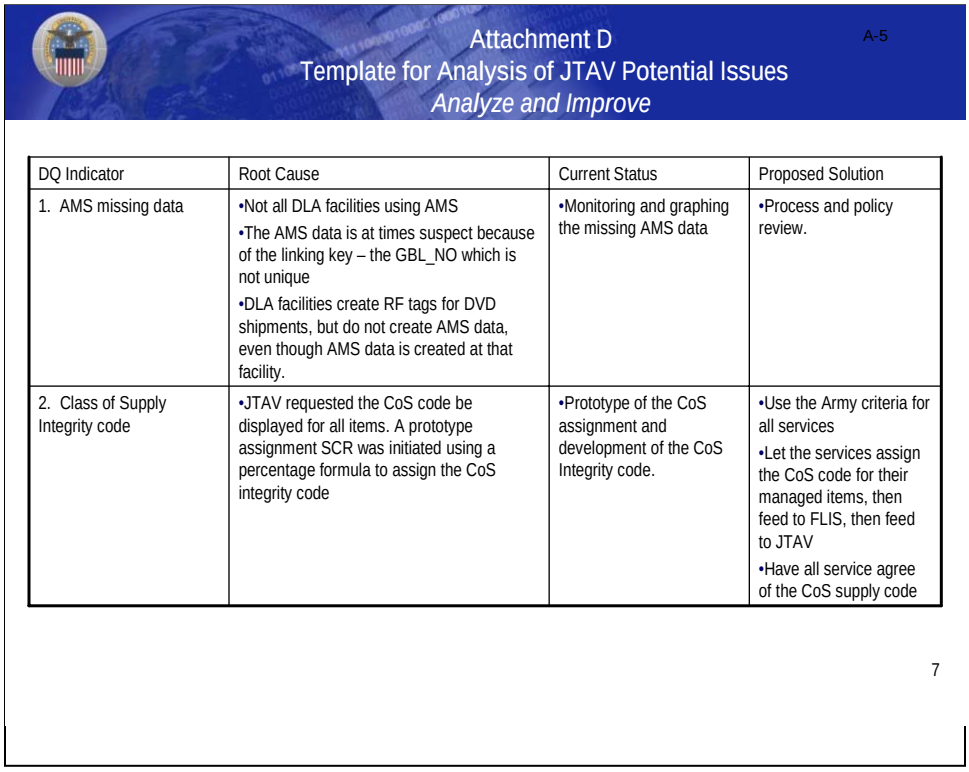

FIGURE 3

STEP 3: IMPROVEMENT

Recommended improvements will identify data quality problems to be solved, *root causes* for data quality problems, and corrective actions necessary to improve the data. The PM will consult with the data supplier and monitor the levels of data quality achieved using the following root cause areas. An improvement plan of action includes, but is not limited to the following:

a. *Internal system error* – errors are resident in the data system automated programming code.
b. *Policy problem* – a failure on the part of workers or managers to comply with one or more policies.
c. *Procedure problem* – a failure on the part of workers to comply with written or implied procedures.
d. *Training problem* – personnel do not understand the correct policy/procedure to accomplish work.
e. *Interface system error* – data error occurring when two or more data systems share data values.
f. *Unassigned error* – all errors that do not fit in above categories.

To execute the recommendation(s) and monitor the implementation, one or more of the following *improvement categories* must be addressed in the System/Product Action Plan:

- Process Improvement: Focus to improve the functional processes that are used to create, maintain and disseminate data.
- System Improvement: Software, hardware, and telecommunication changes can aid in improving data quality.
- Policy and Procedure Improvement: Resolve conflicts in existing policies and procedures and develop appropriate guidance that will institutionalize the behaviors that promote good data quality.
- Data Design Improvement: Improve the overall data design and use DLIS data standards.

The JTAV Action Plan will continue to be documented with the following:

º Root causes
º Desired improvements
º Method and/or tools needed to accomplish improvement
º Improvement timeline/milestones
º Improvement categories
º Presentations provided to the DCB for improvement approval
º DCB responses (approval or denial)
° ROI analysis when dollar values can be identified from suggested improvement

STEP 4: IMPLEMENTATION

The PM will document the System/Product Action Plan with all actions taken to implement the *approved* improvements, including any customer feedback and influencing factors. The JTAV Action Plan will document all actions taken to implement improvement and provide a dated status of each action, customer feedback regarding improvement and factors influencing the implementation of recommendations.

STEP 5: MONITOR/MEASUREMENT

Improvements made to JTAV will be re-measured, monitored for effectiveness and documented on the System/Product Action Plan.

STEP 6: REPORT/AUDIT

The PM will be expected to provide quarterly Product/System Updates and Data Quality Issue Status Updates. The data quality process will be ongoing. The PM and DS will continuously review the system/product and revised the “Data Quality Issues” on an as needed basis. The ‘stop light’ chart below is an example of the reporting process (figure 4).

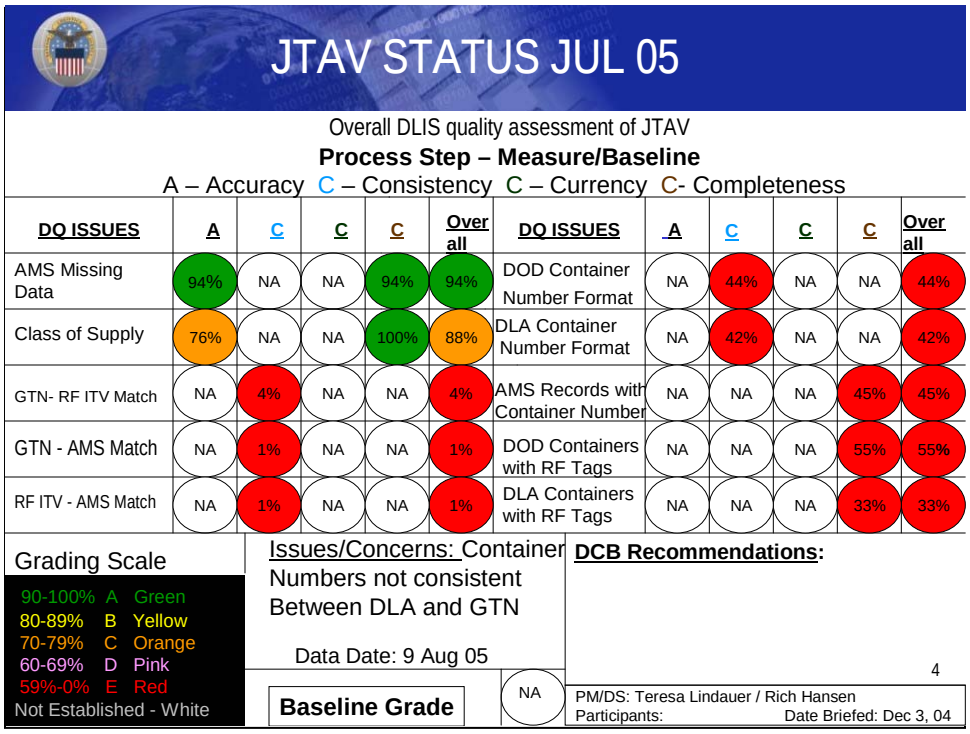

FIGURE 4

SECTION IV. TECHNOLOGY

Technology provides many opportunities for ensuring data quality. This technology can take the form of program system edits or new tools to analyze and detect data issues. Which ever form it takes, technology should be seen as a viable means of enhancing data improvements but not the overall solution.

Industry has acknowledged data quality efforts as a growing business opportunity. More and more emphasis is being placed on developing technology. Current types of technology include:

- data analysis – data and metadata applications to capture specific fields from various systems and compare data
- data reengineering - performs data cleansing based on user-defined business rules
- name/address cleansing – parses specific information as in related elements to ensure standardization and accuracy
- relationship identification - supports matching and merging related data.

DLIS uses two approaches in its data quality effort. One is a contractor who will use ad hoc JTAV queries to document JTAV/AV data quality issues. The second is a commercial off the shelf (COTS) profiling software. The purpose of the profiling tool is to assess the composition, organization, and quality of databases. The profiling tool in no way, shape, or form will change or attempt to change any of the data or operating processes of the systems or databases. We do not want to profile the entire database, rather specific tables.

SECTION V. COST

Improving the data quality of an organization results in both *"tangible"* and *"intangible"* benefits. Intangible benefits are difficult to measure due to the qualitative and subjective nature. These include: improved customer satisfaction, consistency between systems, a decrease in the loss of sales and loss of life. Tangible cost-benefit analysis includes areas such as: the cost of the customer receiving the wrong product due to invalid attributes; the cost of not providing systems and products in a timely manner; the cost of incomplete data, upon which the war fighter bases critical decisions.

VI. CHALLENGES

- DLIS is not always the authoritative source.
- Most existing quality checks are self-certified which poses a conflict of interest and margin of error.
- How to change the mindset and realize that it is ok to admit that there is poor data (identify it and fix it).
- Employees and managers may view data quality as adding to an already heavy workload.
- Organizational buy-in will be crucial for success.
- Training the workforce in a timely and efficient manner.

[1] (DLIS Data Quality Plan)

CONCLUSION

This paper outlines the aspects of the DLIS data quality plan and one of the systems involved in the prototype effort. The plan emphasizes four attributes that are used to measure DLIS data quality; accuracy, consistency, currency, and completeness. The data quality plans six steps encompass defining the environment, establish a baseline measurement, investigate root cause analysis and provide recommended solutions. Implementing an action plan along with monitoring the improvement and measuring the improvement are the next to the last steps. Reporting the status of the system/product to the corporate board will ensure the quality process is moving forward. One goal of the DLIS data quality effort is to document the data quality issues and to assist the PM in the clean up and to ensure the proper edits and safeguards are in place to prohibit any future errors.

DLIS is the program manager for approximately 120 DLA systems; however DLIS does not own the data. The focus of the DLIS data quality effort with the JTAV program office is on DLA owned systems. As mentioned earlier DLIS does not own the data we are the data stewards. A decision was made to concentrate on DLA managed systems (stay in our own backyard). It was decided to use AMS which provides in-the-box (containers and air pallets) visibility for DLA. AMS is a DLA owned system and feeds directly to JTAV/AV.

JTAV takes data from 39 sources and provides a fused view of the data. JTAV provides the Unified Commands the ability to locate assets in storage, in process, in transit, and in theater. JTAV provides users a web access to this information. Additionally, each Command has a JTAV server that supplies data by either reaching back to the source as needed or by storing the data forward in anticipation of users' needs.

A second system is involved in this effort, the GTN system. GTN gives its customers located anywhere in the world a seamless, near-real-time capability to access – and employ – transportation and deployment information. GTN is an automated command and control information system that supports the family of transportation users and providers, both Department of Defense (DoD) and commercial, by providing an integrated system of in-transit visibility information and command and control capabilities. GTN collects and integrates transportation information from selected transportation systems. In keeping with modern technology, GTN is completely available on the Internet's World Wide Web.

The warfighter requires the best possible information and DLIS is postured to lead the charge ahead towards that end.

REFERENCES

1. Defense Logistics Information Service, (2004). *DLIS Data Quality Plan*, Battle Creek, MI, (2-10), Draft 16.

2. Defense Logistics Information Service, (2003). *DLIS Year in Review 2003*, Battle Creek, MI, (3).

3. Global Transportation Network, (2005). Retrieved from the World Wide Web, https://www.gtn.transcom.mil/public/home/aboutGtn/index.jsp This is a secure DoD access only site.

4. Joint Total Asset Visibility, (2005), JTAV Program Management Office, Battle Creek, MI.

5. Science Applications International Corporations, (2004). Joint Total Asset Visibility (JTAV) Program, Data Quality Analysis, Automated Manifest System (AMS).