DETERMINING THE COST AND EFFECTIVENESS OF ENHANCING DATA IN THE U.S. DEFENSE LOGISTICS AGENCY SUPPLY CHAIN

(Research Paper)

L. Robert Pokorny
XSB, Inc.
pokorny@xsb.com

Abstract: We present here a methodology for determining the benefits that can accrue from efforts to enhance data quality relative to the cost of those efforts. The methodology proposes a Data Enhancement Index (DEI) which is a vector sum of measurements of aggregate properties of a legacy data collection along three axes. The first axis measures the current state of the business process that the data collection is meant to support. The second axis measures the current state of the supporting data. The third axis measures the ease with which current data can be enhanced. The construction of the DEI is presented in the context of the U.S. Defense Logistics Agency (DLA) supply chain management and the legacy data that supports it. The DEI is computed for various classes of repair parts that the DLA purchases and indicates which classes of repair parts would achieve the greatest improvement in supply chain performance from an effort to enhance technical data about the parts purchased.

Key Words: Data Quality, Data Enhancement, Quality Measurement, Supply Chain Management, Predictive Indicator of Quality

INTRODUCTION

The Defense Logistics Agency (DLA) is an agency within the U.S. Department of Defense charged with providing worldwide logistics support for the missions of the Military Departments. DLA also provides logistic support to a number of other Federal agencies, foreign governments, and international organizations. One of DLA’s major responsibilities is providing repairable and consumable parts for weapons system maintenance. In connection with this responsibility, DLA maintains strategic and operational data for a supply chain of almost 4 million active parts.

The nature of the DLA supply chain for weapons system parts is unlike a supply chain for a commercial organization because it handles an enormous amount of diverse items, usually purchased and stocked in small quantities with unpredictable demand. To effectively manage such a supply chain the DLA relies on a number of internal databases that describe the physical properties of parts purchased and the logistic properties relating to sourcing, storing, and delivering these parts. The quality of information contained in these databases varies from part to part, and DLA supports a number of programs to improve this quality [4, 6]. However, DLA’s resources are not unlimited and deciding which aspects of data quality are most critical to address becomes a task of significant importance.

In this paper we present a methodology to evaluate the effectiveness of efforts to improve the quality of technical data for categories of supply parts to improve the performance of the DLA supply chains for those part categories. Using this methodology, we are able to suggest which categories of parts are most likely to benefit from an effort to improve the quality of technical information. The following is an outline of the organization for the rest of the paper. The Background section discusses the data landscape for parts managed by the DLA and some quality issues that emerge in this landscape. The Rationale and
Purpose section outlines the specifics of enhancing technical data about parts and our reasoning that led to the methodology presented. In the Methods section we present the details of the methodology and its relation to DLA legacy data. The Results section presents the findings produced by application of the methodology. The Discussion section makes suggestions on how these results could be applied to deciding where the DLA should commit resources to improving technical data quality and how the methodology could be applied to evaluating other data quality efforts. The Limitations section presents potential problems with the methodology. The overall value of the work presented here is discussed in the Conclusion. Finally, an Appendix list definitions for military acronyms used in the paper.

BACKGROUND

To understand the motivation for the work presented here it is useful to understand the nature of data that the DLA maintains about parts that it manages. Items of supply used by the Military are sometimes purchased locally, but, more often they are centrally managed and given National Stock Numbers (NSNs). NSNs are unique identifiers for items of supply and are assigned and maintained by the DLA’s Defense Logistics Information Service (DLIS) in the Federal Logistics Information System (FLIS) [2]. There are about 6 million active NSNs. The DLA manages the supply of about 60% of these, acting as a wholesaler and distributor to the Military Departments.

NSNs are functionally classified into a taxonomy of Federal Supply Groups (FSG), Federal Supply Classes (FSC), and Item Name Codes (INC). Thus, NSN 1560-01-384-6128 is in the FSG 15 - Aircraft and Airframe Structural Components, FSC 1560 - Airframe Structural Components, and INC 29785 - COVER, ACCESS. NSNs are also classified into 10 supply categories that indicate the type of supply chain that supports the item. The 10 categories are listed in Table 1 and the NSN above is in Class IX.

<table>
<thead>
<tr>
<th>Supply Category</th>
<th>Category Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class I</td>
<td>Subsistence</td>
</tr>
<tr>
<td>Class II</td>
<td>Clothing, individual equipment, tentage, tool sets and tool kits, hand tools, administrative and housekeeping supplies and equipment.</td>
</tr>
<tr>
<td>Class III</td>
<td>POL. Petroleum fuels: lubricants, hydraulic and insulating oils, preservatives, liquid and compressed gases, chemical products, coolants, deicers and antifreeze compounds, together with components and additives of such products and coal.</td>
</tr>
<tr>
<td>Class IV</td>
<td>Construction. Construction materials to include installed equipment and all fortification and barrier materials.</td>
</tr>
<tr>
<td>Class V</td>
<td>Ammunition. Ammunition of all types</td>
</tr>
<tr>
<td>Class VI</td>
<td>Personal demand items (nonmilitary sales items).</td>
</tr>
<tr>
<td>Class VII</td>
<td>Major end items. A final combination of end products that is ready for its intended use (principal items); for example, launchers, tanks, mobile machine shops, and vehicles.</td>
</tr>
<tr>
<td>Class VIII</td>
<td>Medical material including medical peculiar repair parts.</td>
</tr>
<tr>
<td>Class IX</td>
<td>Repair parts and components required for maintenance support of all equipment.</td>
</tr>
<tr>
<td>Class X</td>
<td>Materiel to support nonmilitary programs, such as agriculture and economic development, not included in classes I through IX.</td>
</tr>
</tbody>
</table>

Table 1 - NSN Supply Chain Categories
The DLA manages supply of items in a number of these supply chain categories. The supply management problems DLA faces depend on the supply chain category. Items like subsistence, clothing, fuel, and ammunition have steady demand and are bought in bulk usually on long term contracts. Items such as repair parts have unpredictable demand and are usually bought in small lots. It is this second type of item that we address in this paper.

DLA has two main sources of information about the items it purchases. One is the FLIS database that applies to all NSNs. FLIS is the data repository of record for strategic data about NSNs. FLIS data contains, among other things, the FSC and INC associated to the NSN, technical characteristics about the NSN, and supply sources registered to supply the NSN. Technical characteristics are stored as attribute-value pairs where the attribute represents a property of the NSN and the value is the value of that property. The types and number of attributes necessary to describe an NSN are determined by the INC associated to the NSN. Supply sources for an NSN are represented by a reference number which is a supplier-part number pair. The supplier is represented by a Commercial and Government Entity (CAGE) code and the part number is the number assigned by the supplier for the part.

The second source of data the DLA uses is in transition. Traditionally this was the Standard Automated Materiel Management System (SAMMS). SAMMS contains much of the operational data needed for logistics management of items. It contains data on purchase history, inventory levels, back order positions, pricing, demand, and various notes describing constraints on how the item should be purchased. These constraints, among other things, can indicate if the item is open for competition or can only be purchased from approved vendors, what specifications apply to the item being purchased, and whether the government can provide a technical data package providing detailed drawings of the part being purchased. The SAMMS system is now being replaced within DLA by a new ERP system under the Business System Modernization (BSM) program, but the BSM system will contain similar data to SAMMS.

In addition to the data contained in FLIS and SAMMS/BSM, The DLA also has technical data packages (TDPs) for many NSNs which contain detailed drawings and manufacturing specifications. Unlike the data in the FLIS and SAMMS/BSM databases, information in the TDPs is not directly searchable. The information in a TDP is normally provided to a supplier for an NSN after the supplier has been identified. It cannot normally be used to help identify a supplier.

When a Class IX NSN repair item needs to be purchased, the data available to the DLA needs to be consulted to determine from whom it should be purchased. This system works well when a previous supplier of the NSN is used for a new purchase. Problems start to occur when the previous supplier is no longer a viable source for the item. Sometimes alternate supply sources are listed but this is not always the case. Identifying new supply sources for repair items depends on having a good technical description of the item. The technical data available for supporting this description is primarily found in FLIS technical characteristics and the SAMMS Contractor Technical Data File (CTDF) Procurement Item Description (PID). The PID is narrative data containing buyers’ notes about technical issues in purchasing the item. The technical characteristic and PID data can sometimes be supplemented by FLIS reference numbers that associate the NSN to Government and industry specifications and standards. An example of available technical data for an NSN can be seen by looking at the NSN 5331-01-159-2801. The FLIS technical characteristics for this NSN are shown in Table 2. The FLIS reference number information for this NSN is shown in Table 3. The PID narrative data for the NSN is shown in Figure 1.
From technical characteristics we have information on the item name, material and size of the NSN. From reference number data we know that the NSN conforms to Military Standard MS29513 with a size code 370. Finally, the PID note indicates that the NSN is subject to Society of Automotive Engineers specification SAE-AMS-P-5315A. The title of this specification is BUTADIENE - ACRYLONITRILE (NBR) RUBBER FOR FUEL-RESISTANT SEALS 60 TO 70. This indicates the material (NBR Rubber), environment (fuel resistant), and hardness (60-70 Shore A durometer) for the NSN in question. In this case the available data provides sufficient information to identify alternate sources for the part if the current source becomes unavailable.

This is not always the case. An NSN is considered fully described in FLIS technical characteristics if all the relevant properties prescribed for its INC have valid values. Of the approximately 3.5 million active NSNs that DLA manages, 62% are not fully described. In fact, 17% do not have an approved INC, so DLA does not even know what properties are necessary to describe these parts. Furthermore, 71% of DLA managed NSNs have only one supply source listed in FLIS reference numbers and 60% have no data about associated specifications and standards.

Considering the nature of repair parts for weapons systems where demand is difficult to forecast and there is a diminishing base of manufacturers that can meet this demand, DLA expends a lot of resources to identify new supply sources in an environment of incomplete technical data. Many research projects...
within the DLA are specifically designed to address this situation by enhancing technical data. Some have been more successful that others. Having a methodology to assess the potential for success at the start of a data enhancement project would aid the DLA in making wise decisions about the allocation of limited research dollars. The technique described here proposes such a methodology.

**Rationale & Purpose**

XSB, Inc. has been working with the DLA over the last seven years to organize and structure a Coherent View® of part information from legacy data sources. Our efforts have applied semantic data analysis to this legacy data and associated information from commercial Websites to enhance DLA knowledge about the parts they manage. Some of these efforts are described in [7, 9, 10]. Our experience with these projects has generated an awareness that the technical success of the project does not always translate into success for the DLA in improving supply chain management.

As a case in point, consider two recent efforts that we undertook. One attempted to enhance attributes of NSNs that were not fully described by looking for specific values of structured attributes in unstructured narrative DLA data and commercial manufacturer Websites. This effort enhanced attributes for about 100,000 items in a subset of DLA managed NSNs that were commodity hardware items. The second effort assembled a database of information about O-Rings from DLA legacy data, specifications, manufacturer websites, and material scientists at Battelle Labs and provided a website for DLA technicians to access this database. The first effort proved to be very costly for the amount of new information generated. There is a lot of information available about commodity items on the Web, but it is so fragmented that collecting it is manually intensive. Automated collection and extraction techniques had to be retuned for each new fragment of information. The second effort significantly enhanced data for about half of the 38,000 O-Ring NSNs that the DLA manages. It is now being used in pilot programs to identify substitutes and is poised to become a production tool for DLA technicians responsible for sourcing O-Rings.

Based on this experience, we saw the need for a methodology to indicate at the outset of a data enhancement project the likely cost and benefit to be obtained from the successful completion of the project. The methodology needs to address three issues.

1. How much of a problem does the current situation represent? For a specific DLA supply chain project, how poorly is the current supply chain performing.
2. How likely is it that enhanced data will improve the current situation? Is a current DLA supply chain performing poorly because there is a lack of technical data about the parts being purchased?
3. How easy is it to enhance the data? Are there new or additional technical data sources that the DLA could leverage if they were structured and integrated with current technical data?

A methodology that could answer these three questions at the outset of a project using measurements generated from currently available and accessible data would be easy and inexpensive to apply to rank the relative cost and benefit of proposed projects. Using such a methodology, the DLA would be able to make informed decisions about the allocation of limited resources and we would be able to propose new data enhancement efforts that would have the most impact on improving the DLA supply chain. Developing such a methodology is the rationale for the research presented here. The next section describes the details of that methodology.
METHODS

Overall Methodology Description

The methodology we present develops a score for the likelihood of a data enhancement effort to improve the management of a supply chain that depends on the data being enhanced. We call this score the Data Enhancement Index (DEI). The DEI is normalized between 0 and 1 with a 1 indicating a high likelihood that the data enhancement effort will improve supply chain management.

In the context of the DLA there are many supply chains that depend on the DLA technical data about parts. We look specifically at DLA supply chains for Class IX parts, repair parts and components required for maintenance support of equipment. This is a large category that includes everything from landing gear struts for fighter aircraft to filters for air conditioners. To give some structure to this domain we divide all Class IX NSNs by FSC. This groups Class IX NSNs into 282 sub-groups. We then calculate the DEI for each sub-group and rank them on a scale from 0 to 1.

In the following paragraphs we first outline how the DEI is structured and then show how this structure is calculated from available data for the Class IX NSNs

The Structure of the DEI

The DEI is a normalized vector sum of measurements in three orthogonal dimensions that represent three views of data about a group of NSNs. Each dimension is an aggregation of several factors across all NSN records in the group. The three data dimensions can be characterized by the aspects of the data that they measure. The first dimension measures the current state of the supply chain. A high number in this dimension indicates that the supply chain for the group of NSNs is weak. A weak supply chain is one in which current supply sources are unreliable and finding new sources is difficult. The second data measures the current state of data to be enhanced. In the context of our DLA study, this dimension measures how complete technical data is for NSNs in the target group. A high measurement on this axis indicates that current technical data is poor. The third dimension attempts to measure how easy the technical data will be to enhance. A high number here indicates that enhancing data will be easy.

Each of these dimensions is comprised of a number of parameters that can be easily obtained from the legacy data about NSNs. Each dimension is a weighted sum of its parameters that is then normalized to range between 0 and 1. The vector sum of the three dimensions is taken to produce a raw DEI. The idea is that a project to enhance technical data for a group of NSNs represented by a particular FSC would be most likely to have an impact if the current supply chain is weak, the current state of technical data is poor, and the enhancement task is relatively easy. If the supply chain is already functioning well, then enhancing technical data will have little impact. If technical data is relatively complete, then an effort to enhance it will produce little change. If the enhancement process is difficult and time consuming, then the cost of enhancement may outweigh any benefits received. The DEI as a vector sum of these three axes captures this reasoning.

Once a raw DEI is produced, it is scaled to account for the size of the target FSC group and then normalized to a scale between 0 and 1. The scaling factor uses the following formula:

\[ SF = \log(CT_{\text{target}}) / \log(CT_{\text{max}}) \]

Where SF is the scaling factor, CT_{\text{target}} is the count of NSNs in the target FSC group, and CT_{\text{max}} is the count of NSNs in the largest FSC group. Since FSC groups for Class IX NSNs contain anywhere between 1 and 100,000 NSNs, this scaling is used to capture the idea that a data enhancement effort for a large group of NSNs is more likely to yield benefit than the same effort for a smaller group. The raw DEI
is multiplied by the scaling factor to produce the scaled DEI for the FSC group. Finally, the scaled DEI is normalized to between 0 and 1 by dividing the scaled DEI for an FSC group by the largest scaled DEI for all the FSC groups.

**Parameters Composing the Three DEI Dimensions**

In determining parameters with which to measure each DEI dimension we looked for two important criteria.

1. The parameter should be an indication about the dimension being measured
2. The parameter should be easily aggregated from legacy data about NSNs in the FSC group

Using these criteria, the following are the parameter sets for each of the DEI dimensions:

**Supply Chain Weakness Axis**

The following parameters are combined to measure the weakness of the supply chain for NSNs in an FSC group.

1. **Fraction of NSNs in the FSC group that are single sourced** - If an NSN only has a single source or no source indicated in the FLIS reference number data, then this would indicate limited current options on where to obtain it. Even if the current source is reliable, having no identified backup could quickly create problems in managing the supply chain. If a large fraction of the NSNs in the FSC group are in this situation then this is a good indicator of supply chain weakness. This fraction can be easily calculated from available FLIS data.

2. **Average administrative lead time for NSNs in the FSC group** - Administrative lead time is the average time from when the DLA identifies a need to order an NSN to the time when the purchase contract is issued. Long administrative lead times indicate that DLA has difficulty identifying a supplier from which to order. This is again an indicator of supply chain weakness. Administrative lead time is a field in NSN legacy data that can readily be averaged over all NSNs in the FSC group and this average can be compared to the average across all FSC groups.

3. **Average production lead time for NSNs in the FSC group** - Production lead time is the average time from when an order is placed to when the order is received. Long production lead time indicates that the supply source for an item has problems producing the item efficiently. Production lead time is also a field in NSN legacy data and is handled like administrative lead time.

4. **NSN proliferation rate** - This parameter measures the change in the number of NSNs in an FSC group over time. It is derived from three data points, the number of NSNs currently in the FSC group, the number four years ago, and the number eight years ago. These data points are used to determine the rate at which the NSN count is increasing. An increasing NSN count can indicate that too many NSNs are being stocked and that there is possible overlap in the usage of these NSNs. That is two NSNs could actually be the same part used in two different applications. This is again an indicator of supply chain weakness and can be quickly calculated from legacy data.

5. **Fraction of NSNs in the FSC group with non-approved item names** - If an NSN is not assigned to an approved INC it is given the INC code ‘77777’. In this case there is no definition of what properties are necessary to adequately describe the part. If a large number of NSNs in an FSC group have this situation, then it is difficult to identify alternate supply sources for items in this FSC group. This is another indication of supply chain weakness and can be easily determined from the legacy data.

6. **Number of NSNs in the FSC group on backorder** - This parameter looks at NSNs that are backordered. In an FSC group at any given time there are some NSNs that have no demand, there are some NSNs that have demand but inventory and incoming orders can cover that demand, and there are some NSNs that have demand and are in a backorder situation. A large number of backorders is another indication of supply chain weakness. This parameter is calculated by
taking the ratio of NSNs on backorder to the NSNs with positive demand and again can be easily
determined from legacy data

7. *Variation in NSN cost within an FSC group* - Each NSN has a mean acquisition unit cost
(MAUC) in legacy data. This parameter takes the ratio of the standard deviation of the MAUC
for NSNs in and FSC group to the average MAUC for that FSC group. If this ratio is large it
indicates significant variation in price in the FSC group which could result from inefficient
purchasing, another marker of supply chain weakness

**Technical Data Incompleteness Axis:**
The following parameters are combined to measure the incompleteness of technical data for NSNs in an
FSC group.

1. *Fraction of NSNs in the FSC group that are not fully described in FLIS technical characteristics* -
The FLIS data on an NSN includes a field, the Type Item Identifier Code (TIIC), that indicates
whether the NSN is fully described in technical characteristics. This parameter calculates the
fraction of NSNs in an FSC group where the TIIC shows that the NSN is not fully described.
This is an indicator for technical data incompleteness and is easily determined from NSN legacy
data.

2. *Fraction of NSNs in the FSC group where the DLA does not have a technical data package(TDP)* -
The TDP is the drawings and notes necessary to fully describe the manufacturing process for an
NSN. When the DLA has a TDP then it can provide information to any manufacturer on how to
make the part. If a large portion of NSNs in an FSC group do not have TDPs, it is another
indication of technical data incompleteness. The fraction of NSNs in an FSC group having this
property is easily aggregated from NSN legacy data.

3. *Variation in property values for NSNs in an FSC group* - If a given technical characteristic for
NSNs in an FSC group always has the same value, then that technical characteristic is not useful
in discriminating one NSN from another. This parameter measures the variance in values for
properties of NSNs in an FSC group. If this variance is low than the technical data is not very
discriminating. If the variance is high then the technical data is useful in differentiating NSNs in
the FSC group. This is a third measure of technical data incompleteness.

**Technical Data Enhancement Axis:**
The following parameters are combined to measure the ease with which technical data about NSNs can be
enhanced in an FSC group.

1. *Number of different unapproved item names used for NSNs in an FSC group* - When and NSN is
assigned an INC of ‘77777’’ it is not given an approved item name for which specific technical
characteristics are required. In this case the NSN is given an unapproved item name and usually
has a narrative description of the part in technical characteristics. When this occurs it is possible
to enhance technical data by associating the unapproved item name to a similar approved item
name and using the technical characteristics for that approved item name to extract and structure
the information in the narrative description. This parameter is calculated from the count of
different unapproved item names used for NSNs in an FSC group. The more such names, the
more labor intensive it is to enhance technical data from the narrative descriptions. This is one
measure of the ease with which technical data can be enhanced.

2. *Commodity nature of NSNs in an FSC group* - The DLA maintains an electronic portal called
DoD E-Mall [3] for ordering items on-line from commercial vendors. In our work with DLA we
have classified these commercial items to FSC groups. This parameter is derived from the count
of E-Mall commercial items in an FSC group. A large number of commercial items indicates that
the FSC group represents commodity type items and there should be a large amount of
commercial product descriptions on the web with which to enhance technical data on NSNs. This
is a second indicator of the ease with which technical data can be enhanced.
3. **Variation of INCs for NSNs in an FSC group** - The technical characteristics used to describe an NSN with a given INC are defined in a Federal Item Identification Guide (FIIG). Each FIIG defines properties for a related series of INCs in a particular product area and acceptable values for those properties. Enhancing technical information about parts often requires extracting structured property-value pairs from unstructured narrative part descriptions. To do this, we employ automated extractors that are tuned for a particular product domain. These extractors are built at the FIIG level. This parameter counts the number of different FIIG documents that are represented by the INCs for the NSNs in an FSC group. The more different FIIGs represented the more effort is necessary to build and tune different automated extractors that will enhance technical information for the NSNs in the FSC group.

4. **Number of specifications associated to the FSC group** - One approach to enhancing technical information for NSN in an FSC group is to infer properties of the NSNs from standards or specifications that are associated to the NSNs in that group. The Defense Standardization Program Office maintains the Acquisition Streamlining and Standardization Information System (ASSIST) Website [1] that contains information about all current government and industry specification that are adopted by the Department of Defense. Each adopted standard or specification is assigned to a specific FSC. This parameter counts the number of specifications and standards associated to the FSC group. The lower this count the less effort that is required to build inference rules for NSN properties from the relevant standards and specifications. So this is another measure of the ease of enhancing technical data.

5. **Fraction of NSNs in FSC group associated to ASSIST specifications** - This parameter determines the fraction of NSNs in an FSC group that have an associated specification or standard that is assigned to the FSC in ASSIST. This indicates how many NSNs in the FSC group have the possibility of enhancing technical data by inferring properties from specifications. This is another indicator of how easy it is to enhance technical data in the FSC group.

6. **Non-randomness of part numbers in the FSC group** - A third way of enhancing technical data for NSNs in an FSC group is to infer properties from the part numbers associated to the NSNs. Manufacturers often encode values for properties of their parts in the part numbers they assign to parts. A simple example of this would be the battery codes ‘AAA’, ‘AA’, ‘A’, ‘C’, and ‘D’ encoding information about the dimensions of the battery. This parameter is calculated by reviewing the different part numbers associated to each CAGE code for the FLIS reference numbers on all NSN in the FSC group. An algorithm is applied to each such set of part numbers to score how non-random the part numbers are in that set. These scores are then aggregated to give a total score for the non-randomness of part numbers in the FSC group. If part numbers are highly non-random, then it is likely that manufacturers encode part properties in their part numbers. This parameter captures that idea and is another indicator of the ease with which technical data can be enhanced.

The parameters listed above are easily measured from various legacy data and metadata about NSNs. Each parameter is normalized to fall between 0 and 1 so that 1 indicates the strongest measure along each axis; i.e. weakest supply chain, most incomplete technical data, or easiest to enhance technical data. Each parameter is then multiplied by a weighting factor, the weighted products are summed, and the total is divided by the sum of the weighting factors. This produces the final measure between 0 and 1 along each axis. The DEI is computed from these three measures as described above.

**RESULTS**

The DEI described in the last section was calculated for each of the 282 FSC groups that contained Class IX NSNs. The raw scores for each parameter were entered into a spreadsheet and the formulas to calculate the measure along each axis and the raw, scaled, and normalized vector sums of these measures were set up and the final DEI for each FSC group was produced. These were then ranked from highest
DEI to lowest. This calculation was performed with all parameter weighting factors set to 1 so that all parameters were equally important in determining their respective axis measures. A view of the spreadsheet results is presented in Figure 2.

Our goal was to show that the success we had in the data enhancement effort for O-Rings could have been predicted from the DEI. The O-Ring project is ongoing, but preliminary estimates indicate it could produce a four fold return on investment over the next five years through reduced inventory and more economic purchasing. The FSC code for O-Rings is 5331 and Figure 2 shows that the DEI for FSC 5331 ranks it as number 3 out of the 282 FSCs measured. It is also interesting that most of the high ranking FSCs represent commodity type engineered items such as resistors, screws, capacitors, bolts, etc. These types of items are characterized by being well defined in specifications and having standardized part numbering systems from most manufacturers. These are good indications that information is available to enhance technical data. What was somewhat more surprising is that these FSC groups also showed relatively weak supply chains with relatively poor current technical data. Based on this, we feel the DEI can identify other product domains for the DLA where effort to enhance technical data will have a marked effect on supply chain management.

**DISCUSSION**

The goal of this research was to develop a methodology with predictive value to indicate where effort to improve data and information quality would be most likely to produce benefit. The idea behind this methodology is that a collection of information supporting some business process can be viewed as an entity with certain measurable properties and the entity can be partitioned in some way so that comparisons of these properties can be made across the partitions. In the case of the DLA the business process is supply chain management, the collection of information is the technical and logistics data contained in the DLA’s legacy information systems, and the partitions are based on FSC product categories.

The measured properties of the information collection entity are arrayed along three axes. One axis represents how well the current business process is operating. Another axis represents the current state of data that supports the process. The third axis measures the effort necessary to change the current data. Again, in the DLA example presented here, the first axis looked at supply chain weakness, the second looked at incompleteness of technical data about parts, and the third axis looked at the cost of enhancing the technical data.

Data quality frameworks such as that presented by Wang and Strong [8] address the middle axis and do so by defining derived quality dimensions such as accuracy, relevance, and timeliness that require a significant effort to compute from the underlying data. Such frameworks have been shown to give a very good picture of the quality of information in an information collection relative to the purpose for which the information was assembled. They speak less to the cost of improving that quality and the potential savings that might be achieved from that improvement. Gackowski [5] addressed some of the cost-benefit issues related to computing and using different data quality attributes for specific business processes, but this still didn’t obviate the need to compute the relevant data quality attributes.
<table>
<thead>
<tr>
<th>Part Class</th>
<th>Technical Data Weakness Ass.</th>
<th>Supply Chain Weakness Ass.</th>
<th>Data Enhancement Ass.</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSIC Class Name</td>
<td>Precision</td>
<td>Non-Technical</td>
<td>Risk Management</td>
</tr>
<tr>
<td>3D Printing</td>
<td>0.989</td>
<td>0.989</td>
<td>0.989</td>
</tr>
<tr>
<td>Electromechanical Systems</td>
<td>0.999</td>
<td>0.999</td>
<td>0.999</td>
</tr>
<tr>
<td>Electronic Components</td>
<td>0.999</td>
<td>0.999</td>
<td>0.999</td>
</tr>
<tr>
<td>Metal &amp; Plastic Parts</td>
<td>0.999</td>
<td>0.999</td>
<td>0.999</td>
</tr>
<tr>
<td>Wire &amp; Cable Assemblies</td>
<td>0.999</td>
<td>0.999</td>
<td>0.999</td>
</tr>
<tr>
<td>Electrical &amp; Electronic Systems</td>
<td>0.999</td>
<td>0.999</td>
<td>0.999</td>
</tr>
<tr>
<td>Other Fabricated Parts</td>
<td>0.999</td>
<td>0.999</td>
<td>0.999</td>
</tr>
</tbody>
</table>

Figure 2 - The DEI Spreadsheet
The work we present differs from these approaches in two major aspects. First, the properties that are used to compute our data dimensions are derived directly from the legacy data. These are usually aggregates of specific fields in the data such as average lead time or functions that are computed directly from data values such as non-randomness of part numbers. Second, the DEI is computed for partitions of the legacy data so that these partitions can be compared. This is very useful in a large information system where limited resources can be applied to quality improvement.

LIMITATIONS
The major limitation to applying this methodology is identifying the parameters that compose the measures along the three axes. This requires a good understanding the purpose and semantics of the legacy data and metadata. In the case of the DLA, our work over the last seven years has given us this understanding so that we know which data elements are indicative of the qualities we want to measure. Much of this knowledge could be transferred to work for similar supply chain management issues in the commercial world. As the data domain of interest diverges from the area of repair parts the learning curve to apply the methodology becomes steeper. Applying the methodology to a new data domain requires identifying the processes the data supports, how the effectiveness of those processes can be measured from the supporting data, what part of the supporting data influences the effectiveness of the processes, and how can the enhancement of that supporting data be achieved and measured.

A second limitation with the DEI is that data and information it applies to need to be partitioned. With the DLA data a partition on FSC group was an obvious choice. When looking at a large and complex information collection like the DLA legacy data, the purpose of partitioning is to provide a way to compare efforts at data enhancement that only apply to a portion of the data. In other data domains a partitioning scheme that is useful may be harder to identify.

CONCLUSION
We have presented here a methodology for examining large and complex data collections that support ongoing business processes in a way that provides a quick measure of where efforts to enhance data quality would be most effective. We call this measure the Data Enhancement Index. We have shown how this index could be calculated in the context of DLA legacy data supporting their supply chain management processes. The DEI was calculated for different FSC groups of parts that the DLA manages and the resulting numbers agree at least qualitatively with our experience as to which data enhancement efforts provided the most benefit. While the presentation is developed in the context of the DLA, we believe the methodology can be applied to any collection of information that supports a business process.

The methodology looks at a set of data as an information entity. This entity has a context to which it applies, usually a specific business process that the data collection supports. It also has a usage; the quality of data in the collection has a specific relation to the operation of the business process. Finally, there is a measurable cost to improving different aspects of the data in the information entity. Given an information entity with the above characteristics, the methodology provides a measure for determining which data quality enhancement efforts would generate the most benefit relative to improving the performance of the business process supported.

APPENDIX: A GLOSSARY OF MILITARY ACRONYMS
ASSIST - Acquisition Streamlining and Standardization Information System. A Web based system providing data on all Government and Industry standards and specifications adopted by the Department of Defense.
BSM - Business Systems Modernization. An ongoing ERP implementation in the Defense Logistics Agency set to replace the Standard Automated Materiel Management System for maintaining all transactional and logistic data necessary for ordering items of supply.


DLA - Defense Logistics Agency. An independent agency in the Department of Defense tasked with providing logistical support to the Military Departments.

DLIS - Defense Logistics Information Service. An organization within the Defense Logistics Agency that maintains the Federal Catalog data of record on all items of supply.

FIIG - Federal Item Identification Guides. Documents maintained by the Defense Logistics Information Service that describe the technical properties necessary to catalog items of supply.


FSC - Federal Supply Class. A code for a category of items of supply.


INC - Item Name Code. A code for a particular type of item within a Federal Supply Class whose distinguishing characteristics are defined in a Federal Item Identification Guide.

MAUC - Mean Acquisition Unit Cost. The average current cost the Defense Logistics Agency incurs for an item of supply.

NSN - National Stock Number. A unique identifier for a specific item of supply.

PID - Procurement Item Description. A narrative description of specific constraints on purchasing an item of supply. The PID is part of the Contractor Technical Data File.

SAMMS - Standard Automated Materiel Management System. The current system of transactional and logistics data within the Defense Logistics Agency used for ordering items of supply. It is being replaced by the Business System Modernization ERP implementation.

TIIC - Type Item Identification Code. A code in the Federal Logistics Information System indicating if the cataloged technical characteristics for an item of supply are sufficient to describe the item for ordering.

TDP - Technical Data Package. A package of drawings and documents that can be provided to a vendor that will provide all information necessary to produce an item of supply.

REFERENCES
