The Sufficiency of Logistics Data?

"The more you use the data, the better it gets"
"It's better to model with some data than not to model at all."

Whether we have sufficient quality data to underpin logistic support decisions has long been argued with valid questions about its accuracy, consistency, coherence and completeness. The problem is not having too little data but much more than can be managed and assured. Many different sources are used without robust policies and processes to check for quality and quantity which adds to uncertainty. Differing ownership and the commercial value of data impedes creation of a single source of truth. The time and effort needed to collect, collate and cleanse the data for reasonable confidence is large. Lack of confidence in data can erode Support Modelling and Analysis (SM&A) credibility, results are often ignored and, at worst, SM&A is not conducted in lieu of questionable judgement and 'best guesses'.

These observations are largely true. But do they undermine the purpose of Supportability Modelling & Analysis and invalidate its results? To recover confidence in modelling and analysis recommendations, supporting data must be Dependable, Objective and Evidence-based. How can this be achieved?

What Data is Required?

Data is a vital asset, much more than numbers or text as the real value comes from decisions it drives. Performance measurement enables business decisions with less inherent uncertainty and risk. Techniques to exploit and manage data are needed to extract and cleanse raw data, develop useful information, create knowledge and, through simulation and optimization, generate insight to make decisions - to move up the data, information, knowledge & decision (DIKD) staircase. Common Support decisions include: to manage repairable and consumable inventories; estimate Life Cycle Cost and annual operating & Support cost budgets; and to manage fleets for scheduled, unscheduled and upgrade maintenance. Each is different.

Therefore, the first, most fundamental question is to understand business needs from both DIKD ends. What business decisions must be made, what knowledge is needed to make them, what information is required to drive the SM&A process, so what data is needed. And of course, how to govern the process and assure the answers for acceptable confidence.

Data Assurance

Data must be assured to ensure it is accurate, current and valid for the intended use. A UKCeB Joint Information Group report - The Cost of Data Ownership in 2009 reported many anecdotal examples where Project Teams addressed issues individually with isolated initiatives. It highlighted the need for data governance with consistent policies and approaches, and suggested process improvements to address the lack of coherent efforts to manage and improve data quality. Unfortunately, the examples were assessed retrospectively and in isolation with no collated estimate of the impact on the key Support outcomes - availability and cost.

Ideally, formal data quality systems such as ISO 8000 should be used. This international standard focuses on quality systems and master data management to improve data quality and facilitate the exchange of quality data among organizations. The key features are standards and criteria for quality data including accuracy, completeness, and consistency; management of master data essential for business operations; and portability to allow organizations to use data freely across different platforms and applications. These ideals are undoubtedly sensible but not yet universal. Most logistic data was either developed historically or is still not assured.

It will be many years before we can assume high quality data from source and there is a huge legacy of uncertain quality. It is an uncomfortable reality that most of the logistic data for current and future Systems for at least a decade ahead is already corrupt and, unless addressed, will perpetuate poor decision making, creating waste and excess cost. Cleansing legacy data is the only alternative.

Data Cleansing

Up to 80% of current modelling effort can be to capture, collate and cleanse data before analysis can start. While laborious, the data conditioning burden is often overstated as there are mitigations.

First, most of the availability and cost of a system derive from a few Maintenance or Logistically Significant Items (MSI/LSIs). These constitute only a small proportion of the Bill of Materials. For a large, complex aircraft of 150,000 items, less than 1,500 are MSI/LSIs and perhaps only 100 are critical system drivers. Rules of Thumb quoted in the UKCeB Report suggest that ***"50% of the total repair cost is concentrated on the top 5 items, 80% from the top 10. In Avionics, addressing the top 2 items can reduce cost by 20%."*** The main cleansing effort should be on this sub-set.

Second, most Support data does not change over time – the concept of Static and Dynamic data. Static data does not change unless the system design is changed: part numbers, tools, maintenance procedures are all fixed subject to configuration control. Cleansing Static data is a one-off effort. In contrast, Dynamic data will fluctuate even for fixed designs and must be actively managed to maintain its currency. Usage patterns, prices and manpower costs will all change over time while stochastic data such as reliability and maintainability will have inherent variability. Indeed, the variances are as important as the means as the Variance to Mean Ratio is a critical factor in the calculations. Tools and methods must accommodate these factors to assess confidence and inform appropriate risk provisions.

Third, only a few data fields affect Support decisions. A typical LSAR of many Gb has more than 500 data elements in 104 tables but only 40 contain fields that affect Support decisions. Take inventory analysis as an example, there are 6 key data fields: the equipment breakdown structure; fleet usage parameters; demand rate including reliability and non-attributable failures; item prices; repair turnround and production lead times; and administrative and shipping times. Most of the data in a LSAR is needed to execute Support. Examples include text maintenance procedures; tools and support equipment; packaging details; pilferage codes; contract names and addresses. That data is vital but not the focus here.

In sum, the volume of the data to be cleansed can be reduced to manageable proportions.

Data Management Methods

Two specific data management methods greatly ease the practical challenges.

- **Dedicated SM&A Database**. Once data is cleansed, unless the original data sources are also cleansed, errors will perpetuate, and all the good work will be undone when the next update is loaded. Data management mechanisms are needed to protect cleansed data from re-corruption. Because original data sources are usually beyond the control of SM&A specialists and changes will inevitably take time to implement, it is sensible to establish a separate, dedicated cleansed database as the trusted source for SM&A tools. This also has the advantage of being able to hold alternative 'What If' data to permit experimentation without corrupting the master data.
- **Data Quality Attributes**. To facilitate the cleansing process, a very useful mechanism is to tag each instance of every important data field with a Data Quality Attribute (DQA) such as a number between 0-255. DQAs can be set to reflect a combination of completeness, fitness for purpose, accuracy, currency and provenance by user defined rules. DQAs are usually initiated to reflect confidence in their source. After cleansing, DQAs can be uplifted automatically or manually, individually or by class, or by system, sub-system or supplier, to assess current data quality. Raised DQAs for cleansed data prevent over-writing by subsequent automated data uploads. DQAs are also an ideal means to identify and drive focused data quality improvement campaigns.

Tools with these capabilities exist and are already owned, but not used, by MOD.

Data Assurance & Governance

A comprehensive quality approach is needed to drive quality into every stage of the analytical process: into data sources through cleansing; structuring data repositories to protect cleansed data; and using trustworthy, verified and validated analysis. This is specified in the HM Treasury ***Aqua Book - guidance on producing quality analysis for government***.

The Aqua Book outlines a sensible, achievable set of principles that will help ensure that analysis can be trusted to inform good decision making. It sets out principles for analytical quality assurance to help support commissioning and delivery of fit-for-purpose analysis: proportionality of response; assurance throughout development; and Verification and Validation to ensure that SM&A is appropriate and fit for the purpose for which it is being used. Data quality underpins this approach with **RIGOUR** applied for confidence in the results and outputs from modelling and analysis. Quality analysis must be: **R**epeatable; **I**ndependent; **G**rounded in reality; **O**bjective; have understood and managed **U**ncertainty; and the **R**esults should address the initial question robustly. It is particularly important to accept that uncertainty is inherent within the inputs and outputs of any piece of analysis and we must establish how much we can rely upon the analysis for a given problem.

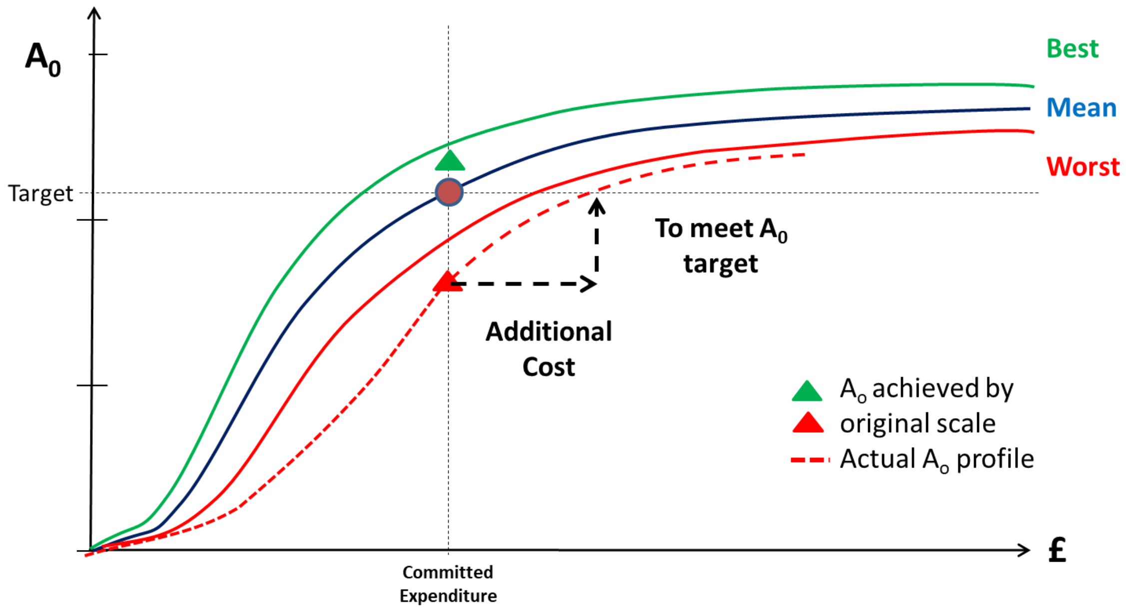
The Use of AI

Some argue that using AI will solve many of these problems. AI can undoubtedly help condition data but since much of it is wrong, incoherent, incomplete and inaccurate and the outcomes have frequently been poor, there is limited access to good data sets for training. We need to guard against generative techniques that generate or extrapolate incorrectly. Besides, once data has been cleansed, existing algorithms and tools are proven, quick to use, and exist without need for expensive development. The use of AI for data conditioning must be treated with caution.

The Impact of Uncertain Data Quality

Uncertain and poor-quality data affects system availability and cost but the impact is rarely quantified. This is an important weakness in current practice but essential for assurance, governance and risk.

Addressing just spares for simplicity, and using baselines and assumptions, inventory optimization identifies the spares need and cost to deliver a target system A_0 as illustrated below. If actual data is different, the achieved A_0 for the same budget will not be quite as good as the Best case but in the Worst case will be far lower. In practice, the achieved A_0 will be within the range of the green and red triangles. Because the Best and Worst cases are not optimized, and the budget has already been spent on incorrect items, more spares must be bought at additional cost to regain the A_0 target.



Adjusting input data to reflect possible differences in data by, for example, increasing price by 10%, reducing reliability by 10-30%, and by lengthening repair cycle times and production lead times by 10%, allows the consequent remedial cost to be evaluated. The cost-benefit of improving data quality can then be quickly and easily evaluated to provide risk-based evidence for the Supportability Case – improve data quality through additional effort, or pay the risk price.

Summary

Inaccurate, incoherent and incomplete poor quality logistics data is endemic for legacy systems and leads to poor Support decisions. While ISO8000 is the long-term solution, it will take decades to work through to deliver reduced Support costs. Data cleansing of key data is the only alternative. AI can help but must be treated with caution since the analytical techniques are quick, easy and proven.

The burden of legacy cleansing need not be as onerous as many suspect provided that the effort is focused on the important sub-sets of items and data fields. Effective data quality protection mechanisms must support the process. To justify the extra time and cost, the potential expenditure to compensate for poor and uncertain data can be evaluated to inform risk budgets.