Chapter 6 – DATA COLLECTION

In terms of time, effort, and resources consumed, collection of data is a major part of a life cycle cost study. Life cycle costing is a data driven process, as the amount, quality and other characteristics of the available data often define what methods and models can be applied, what analyses can be performed, and hence, the results that can be achieved.

It is important to bear in mind that, because the quality and value of life cycle cost and life cycle management analyses is highly dependent on the quality of the available data, good data represent real value for a materiel programme. Conversely, real costs are associated with the collection and storage of data.

This chapter will take the broad view and define data as any type of information used in a life cycle cost analysis.

Data collection for risk analysis is a topic that requires extra attention, and is therefore covered separately in Sub-section 7.4.

6.1 EVOLUTION THROUGH THE LIFE CYCLE

As a system progresses through the life cycle, the types of data available evolve in a number of ways. As this in turn defines the task of data collection and the life cycle costing process in general, it is important to be conscious of these developments.

First and foremost, the amount of data available increases as the system becomes better defined. Obviously, very little is known about the end system when a project begins and all that exists is an identified capability gap or a general concept, whereas, when a system is in service, the system and its environment can be documented in almost infinite detail. Unfortunately, because uncertainty, risks, and opportunities decrease as the life cycle progresses, the need for knowledge is the greatest at the earliest stages. This means that more time and resources should be allocated to the data collection effort at the earlier stages of the life cycle.

The characteristics of the available data will change as well as the progress of the life cycle. Early data will tend to be softer and be in a more aggregated form, because hard numbers and detailed information are not yet available. For example, an early concept may stipulate a high engine reliability, which is later quantified as an expected or required mean time between failures for the engine itself and later still for the various parts of the engine.

Chapter 3 gave an overview of the relevant data sources for each phase of the life cycle of a system. ***It is recommended to anticipate future data requirements and to collect data accordingly.***

6.2 DATA SOURCES

Life cycle costing requires a wide variety of data, and these must be collected from an even wider variety of sources. A distinction can be made between primary data, coming directly from the source, and therefore generally of better quality and utility, and secondary data, which is derived, possibly altered in the process, and therefore of inferior value.

When preparing a cost estimate, estimators should consider all credible data sources. However, primary sources of data should be given the highest priority for use whenever feasible.

Table 6-1 illustrates the difference between primary and secondary data, using the following definitions of primary and secondary sources of data for classification purposes:

- Primary data are obtained from the original source. Primary data are considered the best in quality, and ultimately the most reliable.
- Secondary data are derived (possibly sanitised) from primary data, and are therefore, not obtained directly from the source. Because secondary data are derived (actually changed) from the original data, it may be of lower overall quality and usefulness.

Table 6-1: Sources of Data

Sources of Data	
Typical Data Sources	**Source Type**
Basic Accounting Records	Primary
Cost Reports	Either (Primary or Secondary)
Historical Databases	Either
Functional Specialist	Either
Technical Databases	Either
Other Information Systems	Either
Contracts	Either
Cost Proposals	Either

As already indicated, in the early stages the system itself is non-existent, so data will have to come from comparable systems and programmes. This means that data will primarily come from outside the programme, whereas in the later stages, more data will be generated internally by the project itself. This makes data sources for the earlier phases harder to identify and access and it makes data collection and validation more difficult.

Outside data sources can be industry or other branches of the military or the government. They can be domestic or from other nations. It is worth noting the existence of the NATO agreement on the communication of technical information for defence purposes (Reference: [67]), which covers the communication of proprietary technical information among the signatory NATO nations and NATO bodies.

In these early stages, a certain creativity and flexibility will often be necessary, because most data will have to come from outside the programme. This means that comparable systems and programmes, as well as the relevant data sources within these, must be identified. Data from comparable systems must then be modified based on differences between systems with respect to performance, complexity, maturity of technology, etc. (see Sub-section 6.5 on data normalisation). This task is highly dependent on the specifics of the individual programme, and no generalised or automated methods can be applied. Furthermore, care must be taken to ensure that reference systems are in fact comparable to the system of interest. Use of radically different technology, differences in operating profile, etc., for a new system may mean that data from older systems are irrelevant.

Ideally, it should be proven statistically that a credible relationship exits between the relevant data for a reference system and the system of interest. This, however, is rarely feasible in practice, but ***since the use of assumed or uncertain relationships should be reflected in the uncertainty of the final estimate, it is recommended that evaluation, statistical or otherwise, of the certainty of such relationships are performed.***

At the earliest phases, expert judgment or opinion may be the only available source of information, and it represents the extreme aggregate and soft end of the spectrum. Expert judgment can be viewed as data or as a source of data, but in essence it is a black box analysis of more or less well documented data and experiences. Since it is largely impossible to validate, it is useful if at all possible to get more than one opinion.

Again, these tendencies point to a more manual, creative, labour intensive effort in the early stages of the life cycle.

From the production phase onwards, hard data from tests and in service use, generated within the programme itself, begin to become available. Such data are increasingly collected in ERP (Enterprise Resource Planning) systems and other database systems such as the VAMOSC (Visibility And Management of Operation and Support Costs) databases maintained by various branches of the US military. This in turn allows more automated collection of data in reports and tables for use in the further study of the system itself or of comparable systems at an earlier stage in their life cycle. However, because the need for and availability of various types of data changes throughout the life cycle, data collection is and remains in essence a manual task.

Some data sources, such as contractors and sources external to the defence organisation, will change often, maybe with each new life cycle cost programme, while others, such as maintenance and accounting functions within the defence organisation, will be used repeatedly by the life cycle cost function. It is immensely helpful if there is an understanding of life cycle cost by these repeat providers of data, including how and why it is done and how it is used. This will help ensure that the cost analyst is given the right data of the best quality. Conversely, if there is a widespread feeling that data collection is a pointless chore that generates no value, data quality will drop. This will in turn generate poor estimates, perpetuating a vicious cycle by lending further credence to negative feelings about life cycle costing. Hence, time and effort spent on selling life cycle costing within the defence organisation, informing data providers on how their output is used, may prove very well spent.

It is important in this context to note the value of storing data from past projects and programmes for future use. Though it is important to remember the costs associated with storing data, this just underlines the benefits of anticipating future data needs, and to store the right data in an appropriate format.

6.3 DATA DOCUMENTATION, FORMATS AND STANDARDS

Electronic exchange of data between ERP systems and databases can be a cumbersome and time consuming affair if data formats and data models differ. However, a number of standards exist for the exchange of life cycle data. For some years, the official NATO standard has been the NPDM (NATO Product Data Model), formerly known as NCDM (NATO CALS Data Model). An international standard exists in the form of ISO 10303-239 (STEP, Application protocol 239 – Product Life Cycle Support, PLCS), which has been put forward by NATO to be adopted as STANAG 4661[1].

PLCS provides an application specific but generic and flexible data model for life cycle data. Industry and organisations can tailor this for their specific application using RDL (Reference Data Libraries). Needless to say, this generic data model for all life cycle data is extremely large and complex. This problem is dealt with by defining DEX (Data EXchange Sets), which are subsets of the data model suited for a particular business process, such as the DEX D005 on Maintenance Plan.

Future development of PLCS, including the definition of new DEX, is overseen by a technical committee under the OASIS (Organisation for the Advancement of Structured Information Standards) consortium.

[1] PFP(AC_327)D(2006)0002 (Draft Edition 1), STANAG 4661 on Product Life Cycle Support.

The Cooperation on Defence Implementations of PLCS (CDIP) has been formed to undertake defence related aspects of this work.

As indicated, the PLCS is a very large and technically complex mechanism, and implementing PLCS would be a huge undertaking for any organisation. It is therefore not feasible, for instance, to require industry or other partners to adopt PLCS as a pre-condition for collaboration. In such cases, other, simpler, standards must be found (The OASIS homepage is a good source). Alternatively, ad hoc solutions in the form of agreed upon and documented templates, etc., may be used, but this makes the data harder to use at a later date for other projects or purposes. When possible, officially defined and accepted standards should be preferred.

In the long term, as it has been adopted as an ISO standard and by NATO as a STANAG, and with continued support and development by industry and nations, PLCS has potential to be an important tool to help collect and exchange high quality, well documented data.

6.4 DATA FROM SUPPLIERS/CONTRACTORS

Various mechanisms can be employed to gather data from contractors. The use of a life cycle cost questionnaire is particularly recommended and examples of such questionnaires are provided at Annex C.

Though today data can be shared easier than ever before, it takes time and resources to measure, collect, and manage data. Therefore, it must be understood that data comes at a price, and that it is entirely reasonable for contractors, suppliers, and others to charge a price for data, even if this is part of a major weapons system purchase. The upside of this is that it is then possible and reasonable to make demands on the validity and accuracy of the data received.

In this context, it is particularly important to anticipate future data requirements and to frame agreements and contracts accordingly. The timing of collection and delivery of the data as well as the contents and formats of the required data should be clearly defined. The quality and reliability of data from suppliers is often inferior. Whether this is caused by a lack of incentive or ability, ***it is recommended to have previously agreed upon and well documented templates or standards for the data to be exchanged***.

However, a contractor will often have an interest in presenting a system in the best possible light. Care must therefore be taken to secure that the data received from contractors, or other sources with a vested interest in a programme, are accurate and unbiased. This can take the form of a contract or other binding agreement which puts some form of penalty, possibly some or all of the costs and risks arising from errors and omissions, on the contractor.

The UK MDAL (Master Data and Assumptions List) (see also Sub-section 2.3.2.1; see also Reference [73]) is one well documented mechanism for ensuring that all stakeholders buy into a common and clearly stated understanding of the project and the system of interest. A document like the MDAL needs to be dynamic and iterative, but it also needs to be frozen at certain points (milestones) in the life cycle to provide documentation of decisions taken at this point and support audit of these decisions. This process is illustrated in Figure 6-1.

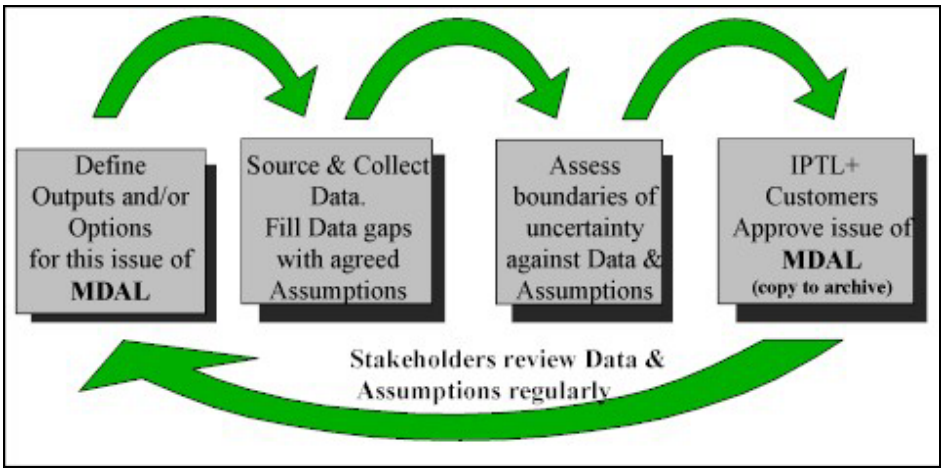

Figure 6-1: Development of Master Data and Assumptions List[2].

A similar function is fulfilled for major acquisitions of the US DoD by the CARD (Cost Analysis Requirements Description) (further details are provided at Section 2.3.2.2). The CARD is used to formally describe an acquisition programme, including the system itself, for the purpose of preparing all cost estimates for the programme. It is provided at major milestone decision points for materiel programmes and for major information system programmes whenever an economic analysis is required.

Furthermore, the US DoD has developed the CCDR (Contractor Cost Data Reporting) system for accumulating actual contractor costs necessary to analyse costs efficiently and effectively. The CCDR contains forms and templates for reporting required cost data and provides extensive guidance on mandatory and recommended policies and processes to be followed by contractors for major materiel systems. The main components of the CCDR system are:

- The **cost and software** data reporting plan. This form, referred to as the CSDR plan, specifies the WBS[3] (work breakdown structure) elements, the specific report format and the reporting frequency.
- The **cost data summary report**. This form captures all contract WBS elements at the level specified in the CSDR plan and includes both recurring and nonrecurring breakouts.
- **Functional cost-hour and progress curve report**. Part I of this form, functional cost-hour report, is directed at selected WBS elements where more detailed cost data are needed. It contains a functional breakout (e.g. engineering and manufacturing) and a cost element breakout (e.g. direct labour and material) within functional categories. Part II, progress curve report, captures recurring costs on lot or unit data for selected WBS elements.

A similar mechanism exists for information systems and other software heavy acquisitions in the form of the SRDR (software resources data report). The CCDR and the SRDR are each described in a separate manual which are available on the internet:

- DoD 5000.4M-1 Contractor Cost Data Reporting (CCDR) Manual, 4/1999 (See: Ref. 78) http://www.dtic.mil/whs/directives/corres/html/50004m1.htm
- DOD 5000.4-M-2 Software Resources Data Report (SRDR) Manual 2/2004 (See: Ref. 79) http://dcarc.pae.osd.mil/srdr/DOD50004M2.pdf

[2] From www.ams.mod.uk/ams/content/docs/wlc/wlcmdal.htm

[3] WBS is similar to Cost Breakdown Structure (CBS) as defined by SAS-028.

While data collection through formal reports such as CCDRs and SRDRs is extremely important and beneficial, there is still no substitute for taking the time to understand and verify the accuracy of historical information, and the programmatic context in which it was obtained.

6.5 DATA NORMALISATION

In Sub-section 6.2, the concept of primary and secondary data was introduced, and it was indicated that primary data are preferable. However, since raw data come from a variety of sources, there is generally a lack of uniformity in data and therefore a certain amount of normalisation will be unavoidable. Generally speaking, data normalisation covers changes and adaptations made to primary data to make it applicable in a given model. It is defined by SCEA (Society of Cost Estimating and Analysis) as:

- To adjust a measured parameter to a value acceptable to an instrument or technique of measurement.
- For a data base: to render constant or to adjust for known differences.
- For cost or dollars/Euros: Then-Year dollars/Euros and/or actuals are escalated to a common Base Year for comparison.

The last definition in particular will be relevant for cost data. However, normalisation can take many different forms and have different specific purposes, such as:

- Adjusting costs to a common year or adjusting to different inflation or discounting mechanisms or other variations in accounting standards.
- Adjusting system or parts costs for technical specifications like size, weight, complexity, technological maturity, etc.
- Adjusting costs or technical performance data such as failure rates for different operating profiles like operating temperature, mileage, etc.
- Adjusting prices for lot size, learning curve considerations, producer capability and maturity, etc.
- Adding cost items not originally included, for example through error or because of a different costing scope, or removing cost items which are not applicable.

Regardless of how data are normalised, exact, complete and detailed documentation of the process is very important. This is the case whether normalisation of primary data is performed as part of the life cycle cost estimating process or secondary data has been obtained for use in life cycle cost estimation. Serious errors can occur if data is not properly understood and interpreted. It is therefore vital to fully understand data and to know where data is coming from.

The next sub-sections describe the relevant issues related to cost data normalisation.

6.5.1 Base Year

The first step is to establish an appropriate base year for data normalisation. A base year is a fiscal year whose mid-point is selected as a reference point for computing an index. A programme base year is usually the year of initial programme funding. Normalising to the programme base year facilitates the analysis of data on a comparative basis during the cost estimating process.

6.5.2 Constant Years versus Current Years

An estimate is said to be in constant dollars or Euros if costs are adjusted so that they reflect the level of prices expressed in the dollars/Euros of a fixed base year. The terms real or constant are used

interchangeably to refer to the purchasing power of the dollar/Euro for the specified base year. When cost estimates are stated in real dollars/Euros, the implicit condition is that the purchasing power of the dollar/Euro will remain unchanged over the time period of the programme being costed. Normalising data due to inflation allows an estimator to track price changes explained by other causes.

Current year costs reflect the purchasing power in existence when expenditures are actually made. Prior costs expressed in current year dollars/Euros are the actual amounts paid out in those years. Future costs stated in current year dollars/Euros are projected amounts to be paid, including the changes in the purchasing power of the dollar/Euro. Terms as current, then-year and nominal dollars/Euro are sometimes used interchangeably.

Cost estimates normally are prepared in constant dollars/Euros to eliminate the distortion that would otherwise be caused by price-level changes. This requires the transformation of historical or actual cost data into constant dollars/Euros. For budgeting purposes, however, the estimate must be expressed in current year dollars/Euros to reflect the programme's projected annual costs by budget appropriation. These annual appropriations actually are expended over a number of years. This requires that the appropriation request takes into account the effect of the anticipated inflation that corresponds to the outlay pattern for each appropriation. The dilemma facing the estimator is how to bridge the gap between the estimate in constant year dollars/Euro and a budget request in current year dollars.

6.5.3 Using Indices

As mentioned before in order to compare cost incurred or estimated for different years asks for converting costs from one year to reflect the price level of another year. Price or other indices can be used to accomplish this conversion. An overview is given (see reference: [79]) on how to select the right index and how to use these indices.

6.5.4 Exchange Rates

The use of foreign exchange rates is a problem unique to analyses performed on international programmes where costs are stated in foreign currencies. This is particularly prevalent in multi-national programmes. It is usually difficult to obtain reliable forecasts of foreign exchange rates. One approach is to assume that if inflation in the foreign country is greater than own country inflation, the rise in foreign prices will be fully offset by currency devaluation (this is the concept of 'purchasing power parity').

In multi-national programmes, usually a base currency and a base year date is agreed upon, e.g. it is agreed upon that all costs will be express in constant Euros, July 1st, 2006 or economic conditions July 1st, 2006. In this case costs calculated in other currencies need to be expressed in Euros using the exchange rate for that particular date. For example, 1 US dollar = 0.83 Euro. This exchange rate would then be used throughout the study.

A typical process for using exchange rates is shown below:

- **Step 1.** If the foreign values are expressed in constant value, note the base year. If they are first expressed in current value, deflate by using the appropriate foreign compound index. The result of this step is that the costs will be expressed in constant currency for a known base year.
- **Step 2.** Multiply the result from step (1) by the own/foreign currency exchange rate for the known base year. The result of this step is the constant currency costs.
- **Step 3.** With the costs now established, multiply these costs by the proper inflation values using the base year established at step (1). These will be the costs to be included in the estimate.

6.5.5 Inflation

Inflation is a consistent rise in the costs (prices) of goods and services over time. To introduce the effect of inflation into economic analysis, the following inflation-related terms are defined:

- Constant currency (Euros, Dollars, GB Pounds, etc.). Constant year currency values are the result of having the effects of inflation removed. Constant year currency values are always associated with a base year; for example, fiscal year (FY) 1995.
- An estimate is in constant currency if the costs for all the work is adjusted so that they reflect the base year level of prices. When prior or future costs are in constant currency, the figures given are adjusted to presume that the "buying power" of the currency was the same and will continue to remain the same as in the base year. The use of constant currency assists in the evaluation of resource requirements over time because it removes distortions which are attributable only to price level changes. With the removal of inflation, the true cost growth of a system can be more readily determined.
- Current or then year currency. Current year values are expressed in the value of the year in which a cost is expected to occur, and therefore reflect the effects of inflation. The term "current year" means that the amount is appropriate for the year in which the money is expected to be expended. When prior costs are stated in current year values, these values are the actual amounts paid out. When future costs are stated in current values, the figures given are the actual amounts which will be obligated including any amount estimated for future price change. When making estimates for the future, assume a base buying power for each currency unit (constant values) and then apply an inflation factor that converts the estimate into current year values.

6.6 REFERENCES

[73] See http://www.ispa-cost.org/PEIWeb/newbook.htm to obtain a copy of the "Parametric Estimating Handbook" from the International Society of Parametric Analysts.

[74] See http://ceh.nasa.gov/ to obtain a copy of NASA's "Cost Estimating Handbook".

[75] PFP(ACQPR)D(2003)2, NATO Agreement on the Communication of Technical Information for Defence Purposes.

[76] http://www.usmcvamosc.com

[77] http://www.navyvamosc.com

[78] ISO 10303-239 (PLCS).

[79] http://www.plcs-resources.org

[80] http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=plcs

[81] http://www.ams.mod.uk/ams/content/docs/wlc/wlcmdal.htm

[82] DoD 5000.4M-1 Contractor Cost Data Reporting (CCDR) Manual.

[83] DoD 5000.4M-2 Software Resources Data Report (SRDR) Manual.

[84] http://dcarc.pae.osd.mil (US Defense Cost and Resource Center (DCARC). CCDR and SRDR manuals are found here).

[85] FAA Life Cycle Cost Estimating Handbook, Investment Cost Analysis Branch, ASD-410, June 3, 2002, Chapter 5.

[86] SCEA glossary.

[87] Cost Estimating Guide, US Department of Energy, DOE G 430.1-1, March 1997.