Data Collection and Normalization

How to get the data and how to normalize it

"The Government are very keen on amassing statistics. They collect them, add them, raise them to the nth power; take the cube root and prepare wonderful diagrams. But you must never forget that every one of these figures comes in the first instance from the village watchman who puts down what he damn well pleases."
-Josiah Charles Stamp, 1st Baron Stamp, Bt, GCB, GBE, FBA

http://en.wikipedia.org/wiki/Josiah_Stamp,_1st_Baron_Stamp

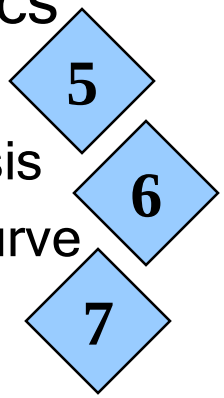
Acknowledgments

- ICEAA is indebted to TASC, Inc., for the development and maintenance of the Cost Estimating Body of Knowledge (CEBoK®)
  - ICEAA is also indebted to Technomics, Inc., for the independent review and maintenance of CEBoK®
- ICEAA is also indebted to the following individuals who have made significant contributions to the development, review, and maintenance of CostPROF and CEBoK®
- Module 4 Data Collection and Normalization
  - Lead authors: Donald S. Ellis, R. Alex Wekluk, Michael T. Yanavitch
  - Assistant authors: Charles E. Casserly
  - Senior reviewers: Richard L. Coleman, Jeff Cherwonik, Fred K. Blackburn
  - Reviewers: Richard B. Collins II, Jason Dechoretz, Peter Andrejev, Angela E. Camp
  - Managing editor: Peter J. Braxton

Unit Index

Data Collection Overview

- Key Ideas
  - Cost as empirical measurement
  - Idiosyncratic accounting
  - Normalization to create "apples-to-apples" costs
  - Painful but vitally important process
- Practical Applications
  - CER Development
  - Cost and Technical Database Development
- Analytical Constructs
  - Transformations
  - Reversible operations
- Related Topics
  - Inflation
  - Data Analysis
  - Learning Curve

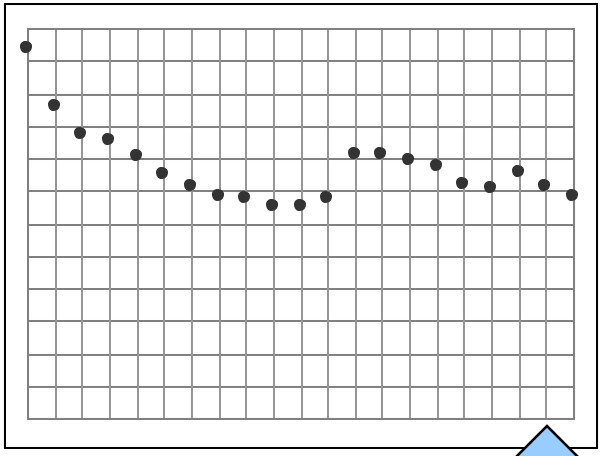


Data Collection and Normalization Within The Cost Estimating Framework

v1.2

Past

Understanding your historical data

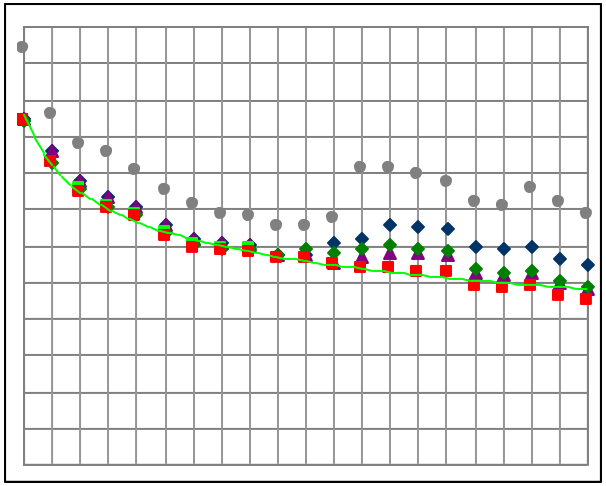


7

Raw data, un-normalized

Present

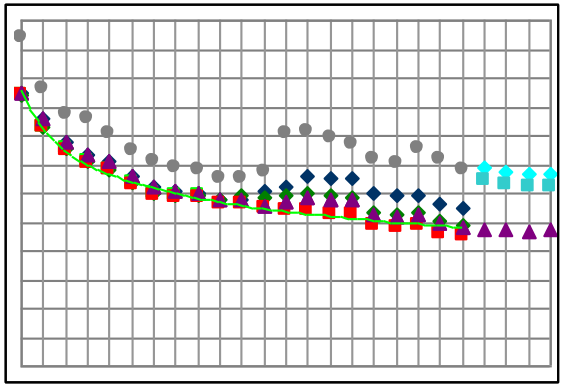
Developing estimating tools



Defensible and traceable normalization

Future

Estimating the new system



Normalization steps reversed

Data Collection Outline

- Core Knowledge
  - Importance of Data
  - Types of Data
  - Considerations
  - Data Collection Process
  - Sources of Data
  - Data Normalization
- Summary
- Resources
- Related and Advanced Topics

Importance of Data

- Historical data are the backbone of a good estimate
  - They provide credibility, accuracy, and defensibility
- Cost analysts must be able to discern data quality
- Data collection is a top priority for cost estimators
  - Not necessarily so for resource managers
  - Need to put forth maximum effort to obtain good data ("beg, borrow, and steal")

1

"Shrimpin' is tough!"
Tom Hanks as Forrest Gump,
Forrest Gump

Data collection and methods development require significant investment

Data Key Principles

- Estimates must be consistent with cost collection methods
- Don't just collect cost data
  - Schedule, technical, performance, programmatic, and operational
  - Risk data for all of the above
    - Initial estimates and final values (actuals)

- Contextual completeness is crucial
- Any normalization step you take with historical data, you must be able to re-apply in estimating
  - Whatever you take out, you gotta put back in!
- Normalization is essentially an estimating step, so must be defensible
  - Never "assume away" cost
- Data collection, normalization, and analysis are iterative
  - Make them part of "business as usual"
- Large, homogeneous data sets are preferred

Aircraft Example: Introduction

1

- Case Study: Estimate commercial aircraft
- Data Sources
  - Manufacturers A and B
  - Federal Aviation Administration (FAA)
- Possible Factors Driving Cost
  - Number of seats
  - Dry weight
  - Composite materials
  - Amount of automation
- WBS from MIL-HDBK-881A
  - MIL-HDBK-881A is superseded by MIL-STD-881C

1.1	Air Vehicle (AV)
1.1.1	Airframe
1.1.2	Propulsion
1.1.3	AV Applications Software
1.1.4	AV System Software
1.1.5	Communications/Identification
1.1.6	Navigation/Guidance
1.1.7	Central Computer
1.1.8	Fire Control
1.1.9	Data Display and Controls
1.1.10	Survivability
1.1.11	Reconnaissance
1.1.12	Automatic Flight Control
1.1.13	Central Integrated Checkout
1.1.14	Antisubmarine Warfare
1.1.15	Armament
1.1.16	Weapons Delivery
1.1.17	Auxiliary Equipment

Types of Data

- Cost, Technical, and Programmatic
- Primary versus Secondary
- Data Quality of Cost Types
- Quantitative versus Qualitative
- Objective versus Subjective

Data – Cost, Technical, Programmatic

5

11

- Cost Data
  - Represents costs (in a specified type of dollars) and hours associated with activities (labor) or materials
  - Cost (or price) is the primary issue, so focus on this
  - Be sure to note cost or price, and what's included (e.g., direct, OH, G&A, fee)
- Technical Data

14

  - Derived from requirements or physical characteristics of systems, may drive cost
    - Ex: Size, weight, MTBF, MTTR, R&M, special security requirements, payload requirements, etc.

15

  - Includes performance and operational parameters (speed, crew composition, activity rates, deployment plans, etc.)
- Programmatic Data
  - Program parameters that explain and drive cost
    - Ex: Schedule, quantity, value chain structure, multi-year procurement, contract type, sole source vs. competitive acquisition, etc.

Performance and Operational Data (subset of Technical) and Schedule Data (subset of Programmatic) should also be collected

Data - Primary vs. Secondary

- Primary Data
  - Obtained from Original Source
  - Unaltered or unchanged
  - Most defensible, best quality
  - Ex: BOM, test results, actual man-hours

4

- Secondary Data
  - Derived from Primary Source
  - Changed or sanitized
  - Lesser overall quality, good for cross-check
  - Ex: Factors, studies, other estimates

8

Data – Quality of Data Types

- Actuals for completed programs/contracts

  are better than

- Estimates at Complete (EACs) for contracts greater than 90% complete

  are better than

- Contract line item prices from Section B

  are better than

- Historical budget data

16

14

Data – Quantitative vs. Qualitative

- Quantitative Data
  - Ordinal – order only
  - Interval – relative scaling
    - Celsius temperature, e.g.
  - Ratio – absolute scaling
    - Kelvin temperature, e.g.
  - Usually Objective, but can be Subjective
- Qualitative Data
  - Categorical (nominal)
  - Often Subjective, but can be Objective

"Nominal, ordinal, interval, and ratio typologies are misleading," P.F. Velleman and L. Wilkinson, *The American Statistician*, 1993, 47(1), 65-72

Data – Objective vs. Subjective

Tip: To be considered objective, similar scores must be obtained in independent trials

- Objective Data
  - Usually Quantitative (e.g., costs, weights)
    - Could be Qualitative (e.g., U.S. vs. Soviet fighters)
  - Collected through formal data collection process
  - Ex: Staff hours, SLOC, Function Points, End Items

12

- Subjective Data
  - Based on judgment – a "feeling" or "understanding" of a characteristic or condition
  - Can be Qualitative (e.g., High Risk) or Quantitative (e.g., 30% more complex)
  - Ex: Complexity, Level of Difficulty, TRL

2

6

Look to gather data from all credible sources, but **Primary**, **Quantitative**, and **Objective** data are best.

Aircraft Example: Data Types

- Data Types

Data point	Primary	Secondary	Objective	Subjective
1	Mfr A	FAA Study	Dry weight	Complexity
2	Mfr B	Internet search	Max. speed	Construction Quality

- What to Get:
  - Cost (Resource) Data
    - Labor hours/component
    - Ratio of development to production
  - Technical Data
    - Weight, speed, power requirements, fuel consumption
  - Programmatic Data
    - Schedule, contract type

Good Contextual Completeness

Quantity and specific model plus accessories are identified

5

- Procurement of one 1998 Dodge Ram 1500 regular cab ST 4x4 truck in standard configuration with 40/60 split cloth seats, optional power package, optional tow package and manual transmission for $18,500 in March 1998 excluding shipping and destination costs ($750).

Costs identified with year in which spent.

Good Contextual Completeness

Specific activity and manufacturer is identified.

- To fabricate a specific sub-assembly, "XYZ" manufacturer required
  - 99.5 labor hours to complete the 1st unit
  - 5th unit required 89 labor hours
  - 10th unit required 84 hours
  - Material costs were constant at $2350 (FY01$) per assembly

Labor hours identified with the production unit number.

What "easy answer" can you get from the second and third data points?

Unit material costs shown with year in which spent.

Poor Contextual Completeness

What are the utilities for? Where? For what period of time? 1 yr? 1 mo?

Year of dollars? BY00? Then Year?

- O&M utilities cost $36,500.

Which utilities? Gas? Electric? Water? Internet?

- Navy ships consume 50,000 barrels of fuel per day

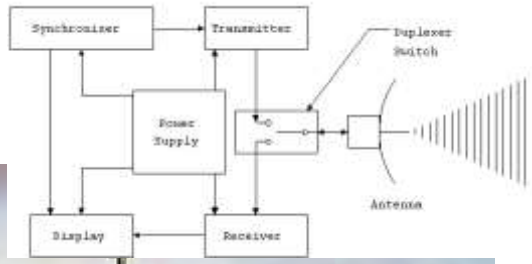
What kind of ships? Only for ships that are underway?

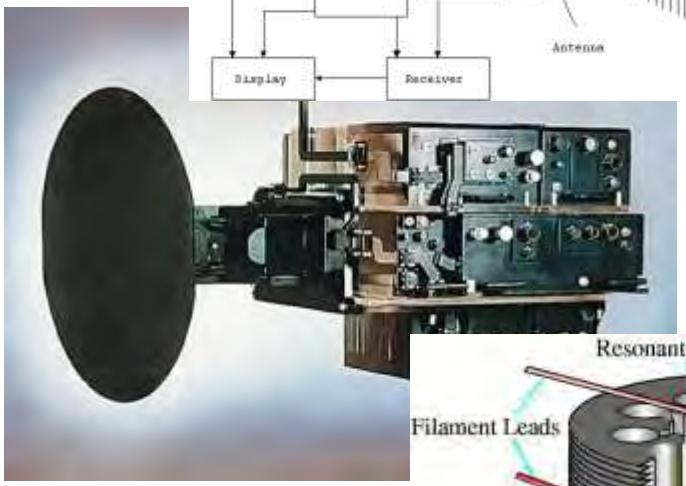
What kind of fuel? JP5? Does this include embarked aircraft?

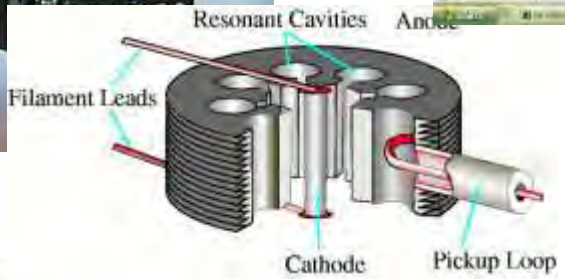
Technical Data Examples

WBS No.	Unit Description	Units / System	Spares	No. Units Produced	Unit Weight (Lbs)	% Weight Margin	Wgt / System	$/Lb	AUC	Dev. Cost	Prod. Cost
1.04	**Mechanically Scanning Pulse Radar**										
1.04.01	Transmitter										
1.04.02	Modulation Circuitry										
1.04.03	Timing Circuitry										
1.04.04	Duplexor										
1.04.05	Receiver										
1.04.06	Traveling Wave Tube										
1.04.07	Beam Steering Unit										
1.04.08	Coaxial Cables										
1.04.09	Signal Processor										
1.04.10	Display Unit										
1.04.11	Antenna Drive Mechanism										
1.04.12	Beam Steering Mechanism										
1.04.13	Slotted Waveguide Antenna										
1.04.13.01	Parabolic reflector										
1.04.13.02	Dipole transmit/receive unit										
1.04.13.03	Horizontally aligned directors										

Programmatic Information	
Organization	
Program Name	
Highest Classification Level	
Office Phone	
Program Manager	
Office Phone	

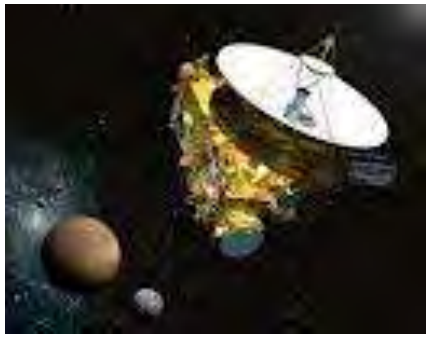


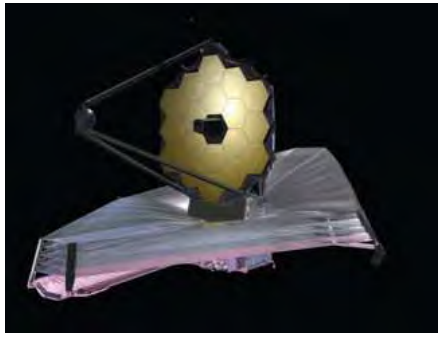




Technical Data Examples (NASA)

NASA Technical Data





Cost Data Example

2

Raw Cost Data	Direct Hours	Direct $	OHD $	Material $	Subc $	G&A $	Total $
WBS X	13150	1,320,250	1,980,375	75,000	160,000	396,075	3,931,700
Cost account X1	5150	440,250	660,375	50,000	150,000	132,075	1,432,700
Job number X1A	1500	150,000	225,000			45,000	420,000
Job number X1B	1650	140,250	210,375			42,075	392,700
Job number X1C	2000	150,000	225,000	50,000	150,000	45,000	620,000
…							
Cost account X2	8000	880,000	1,320,000	25,000	10,000	264,000	2,499,000
Job number X2A							
Job number X2B							
Job number X2C							
WBS Y							
Cost account Y1							
Job number Y1A							

1

11

SECURITY CLASSIFICATION: Unclassified

COST DATA SUMMARY REPORT

Form Approved OMB No. 0704-0188

The public reporting burden for this collection of information is estimated to average 8 hours per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to Department of Defense, Washington Headquarters Services, Executive Services Directorate (0704-0188). Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.
PLEASE DO NOT RETURN YOUR COMPLETED FORM TO THE ABOVE ORGANIZATION.

1. MAJOR PROGRAM	a. NAME:	2. PRIME MISSION PRODUCT	3. REPORTING ORGANIZATION TYPE	4. NAME/ADDRESS (Include ZIP Code)		5. APPROVED PLAN NUMBER
b. PHASE/MILESTONE: Pre-A, B, C-FRP, A, C-LRIP, O&S			X PRIME / ASSOCIATE CONTRACTOR; DIRECT-REPORTING SUBCONTRACTOR; GOVERNMENT	a. PERFORMING ORGANIZATION	b. DIVISION	

6. CUSTOMER (Direct-reporting subcontractor use only)	7. CONTRACT TYPE	8. CONTRACT PRICE	9. CONTRACT CEILING	10. TYPE ACTION				
	CPAF/CPIF			a. CONTRACT NO.:	97-K246800-000	c. SOLICITATION NO.:		e. TASK ORDER/DELIVERY ORDER/LOT NO.:
				b. LATEST MODIFICATION:		d. NAME:		

11. PERIOD OF PERFORMANCE	12. APPROPRIATION	13. REPORT CYCLE	14. SUBMISSION NUMBER	15. RESUBMISSION NUMBER	16. REPORT AS OF (YYYYMMDD)
a. START DATE (YYYYMMDD):	X RDT&E; PROCUREMENT; O&M	INITIAL; INTERIM; FINAL			
b. END DATE (YYYYMMDD):					

17. NAME (Last, First, Middle Initial)	18. DEPARTMENT	19. TELEPHONE NUMBER (Include Area Code)	20. EMAIL ADDRESS	21. DATE PREPARED (YYYYMMDD)
John Doe				

WBS ELEMENT CODE	WBS REPORTING ELEMENTS	NUMBER OF UNITS TO DATE	COSTS INCURRED TO DATE (thousands of U.S. Dollars)			NUMBER OF UNITS AT COMPLETION	COSTS INCURRED AT COMPLETION (thousands of U.S. Dollars)		
			NONRECURRING	RECURRING	TOTAL		NONRECURRING	RECURRING	TOTAL
A	B	C	D	E	F	G	H	I	J
X1	Cost Account X1	x	1382700	$ 50,000	$ 1,432,700				$1,423,700
X1A	Job number X1A	x	420000	$ -	$ 420,000				$420,000
X1B	Job number X1B	x	392700	$ -	$ 392,700				$392,700
X1C	Job number X1C	x	570000	$ 50,000	$ 620,000				$620,000
X2	Cost Account X2	x	2474000	$ 25,000	$ 2,499,000				$2,499,000
X2A	Job number X2A	x							
X2B	Job number X2B	x							
X2C	Job number X2C	x							
Y1	Cost Account Y2	x							
Y1A	Job number Y1A	x							

3

Time Phased Data

	CY95	CY96	CY97	CY98	CY99	CY00	Total $
Program Total							
WBS X							
Cost account X1	143,270	286,540	429,810	429,810	71,635	71,635	1,432,700
…							
WBS Y							
Cost account Y1							

Form 1921
CCDR

Cost Data Example (NASA)

1. Cost Incurred/Hours Worked - Cost and hour data reported in the categories specified in the contract.

2. Estimated Cost/Hours to Complete – Shows the current estimates for performing authorized work included in the most recently executed contract modification, plus additional authorized work

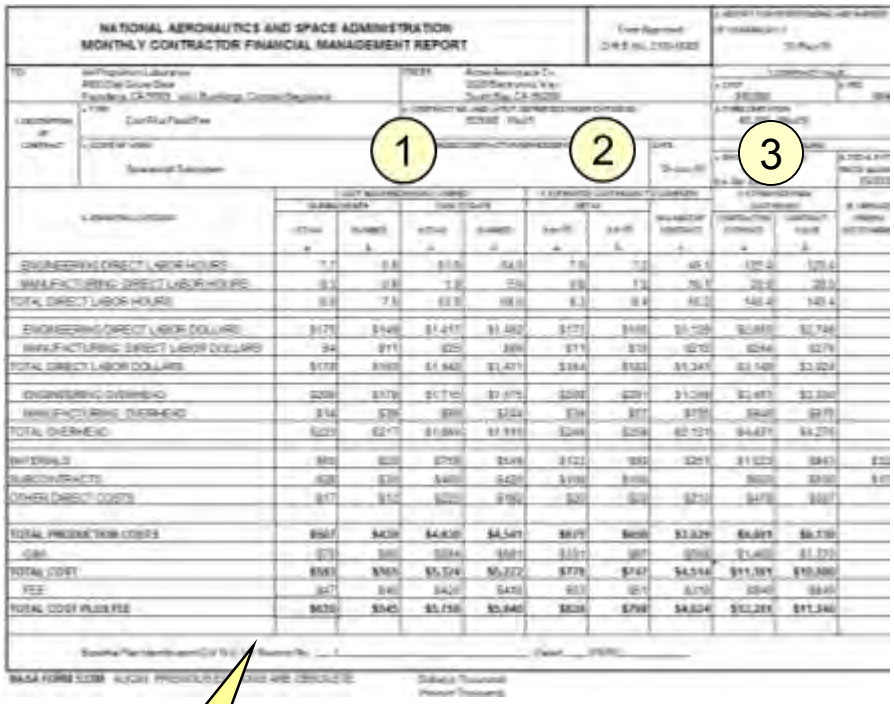
3. Estimated Final Cost/Hours

4. Breakout of Project WBS/CES Structure

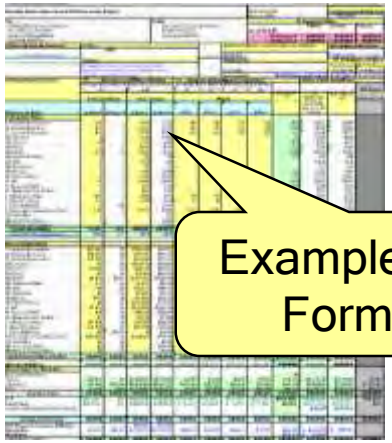
Engineering DL Hours
Manufacturing DL Hours

Engineering Dollar
Manufacturing Dollars

Overhead
Subcontracts
Other Direct Costs
G&A
Fee
TOTAL



Sample NASA Form 533



Example NASA Form 533

18

Considerations

- Availability
- Accessibility
- Validity
- Constraints
- Pitfalls

Considerations – Availability/Accessibility

- Availability - does it exist?
  - Never collected
  - Insufficient granularity or detail
    - Ex: Have summary-level cost, but need sub-element breakout
  - Partial or incomplete
    - Ex: Have costs for the individual integrated circuits and components that go onto the motherboard, but not for the motherboard itself
  - Wrong "format"
    - Ex: Cost for an activity is in dollars, but need it in hours
- Accessibility - can I get to it?
  - Data exists but can not be used
    - Security classification
    - Competition sensitive
    - Proprietary use – may need non-disclosure agreement (NDA)
    - International Trafficking in Arms Regulation (ITAR) controlled
  - Electronic or hard copy?

Warning: "Data denial" is a common defensive tactic

Considerations - Validity

2

13

- Relevancy / Currency
  - Recent enough
- Applicability
  - Comparable program/system
- Cross-check
  - Arrive at similar solution from multiple data sources
- Sanity check
  - Does it make sense? Does is pass the "reasonable person test"?
  - Compare to rules of thumb and experience

Warning: These issues are often used to disable analysis

Considerations – Constraints

3

- Time Constraints
  - Data gathering is time intensive
  - Collection efforts are typically ongoing
- Estimate Requirements
  - What is the purpose or intended use of estimate?
  - What is the purpose of the specific data source?
  - Does data require normalization?

Considerations – Pitfalls and Special Data Needs

- Pitfalls
  - Data definitions vary
    - Different sources - different descriptions
    - Same sources - definitions change over time
  - Processes are dynamic
  - Businesses consolidate or divest
- Special data needs
  - New technology/materials
  - Aggressive schedule
  - Manufacturing process changes
  - Cost reduction initiatives

2

17

9

11

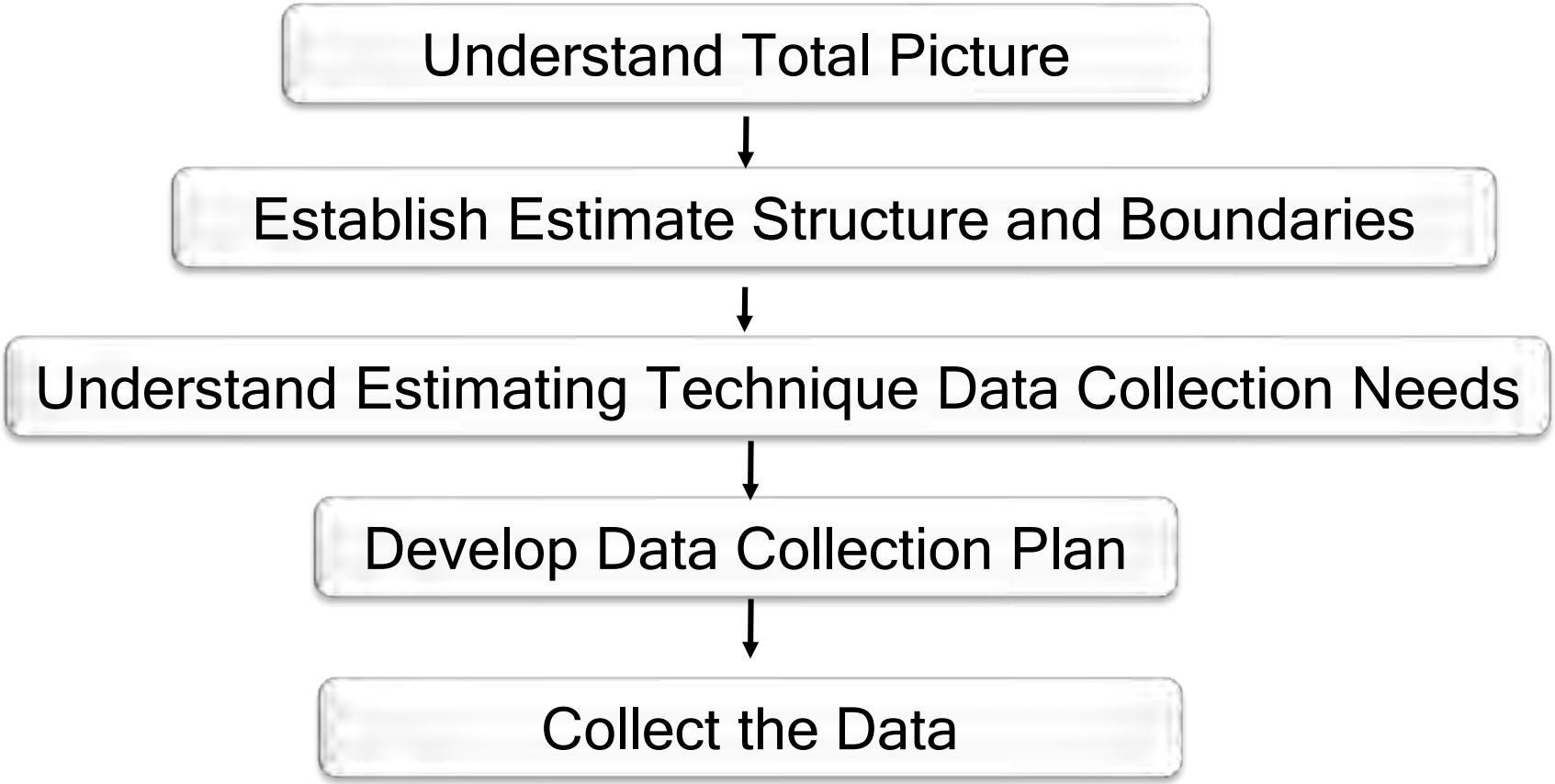
16

Aircraft Example: Considerations

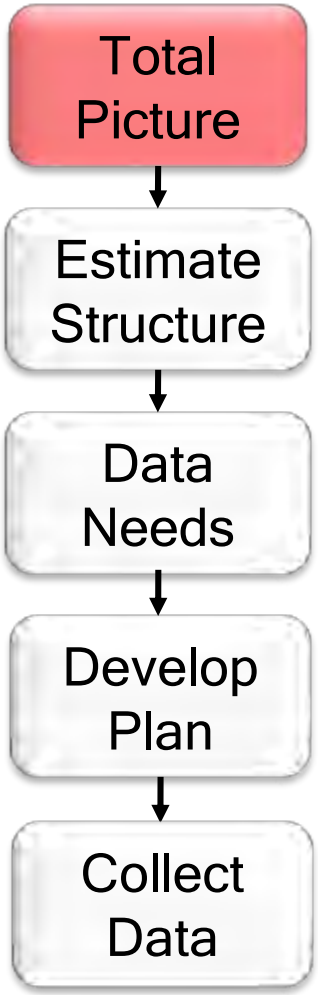
- Availability
  - Non-standard accounting system
  - No development/production split
- Accessibility
  - Manufacturer A may not give the data
- Relevancy / Currency
  - Most available data are from the 1950s – Relevant?
- Time Constraints
  - Need estimate to support budget decisions *now*
- Estimate Requirements
  - No idea how estimate will be used – What to collect?

Data Collection Process

- Cost Analysts must:

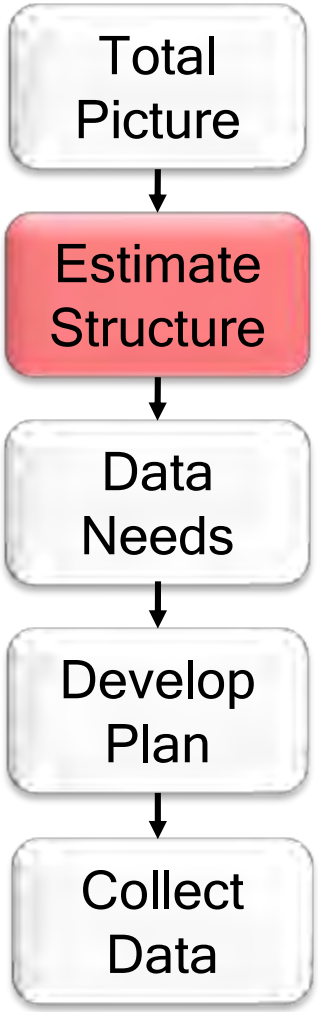


Data Collection Process – Understand Total Picture



- What is the intended use of the estimate?
  - Milestone Review – ICE
  - Budget submission
  - Analysis of alternatives
  - Economic analysis
  - Risk estimate
  - Source selection
  - Earned Value Management (EVM)
- What is the scope of the estimate?
  - Content
  - Time period
  - Organizational scope

Data Collection Process - Estimate Structure



- Establish a Cost Element Structure (CES)
  - Acquisition or O&S or Life Cycle
  - Work Breakdown Structure (WBS)
- Establish boundaries
- Establish sub-breakouts

Tip: CES must be inclusive of *all* costs relevant to the estimate over the appropriate time frame.

Data Collection Process – Data Needs

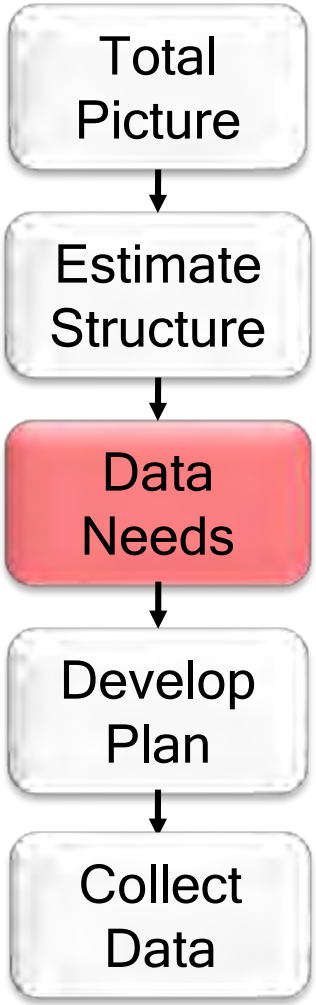
Total Picture → Estimate Structure → **Data Needs** → Develop Plan → Collect Data

- Identify proposed estimating technique(s)
  - List primary and secondary techniques by CES
  - Investigate input data requirements and limitations of candidate models or CERs
- Identify potential data sources
  - Applicable CES
  - Support estimating technique(s)
  - Similar and precedent systems (i.e., relevant historical data points)
  - Understand data limitations
    - Number of data points (n)
    - Proprietary or Competition Sensitive
    - Security Classification

6

Warning: Don't buy the "blank sheet of paper" theory…

Data Collection Process - Data Needs for CER Development



- Factors that drive or influence costs (independent variables), such as
  - Structure weight and material mix
  - Flying hours for aircraft O&S
  - Linear feet and power for cabling cost
  - Operating environment
- How to identify cost drivers
  - Talk to subject matter experts
  - Understand technical and operational parameters
  - Note: Later, scatter plot and statistical analysis will confirm or invalidate choice of cost drivers

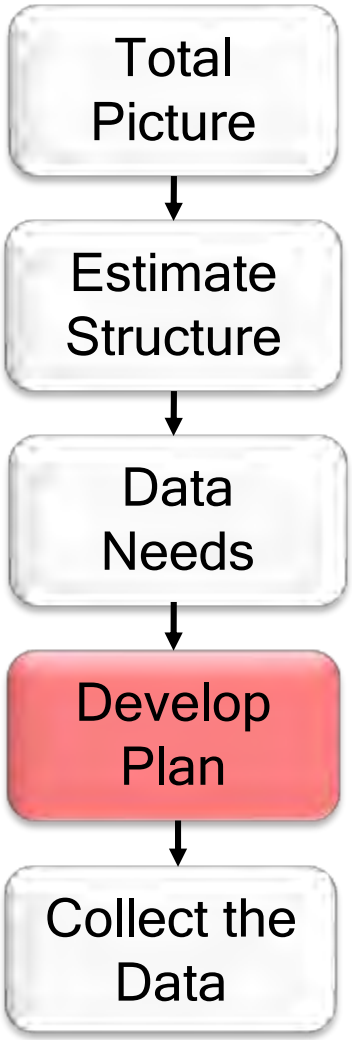
3

6

8

10

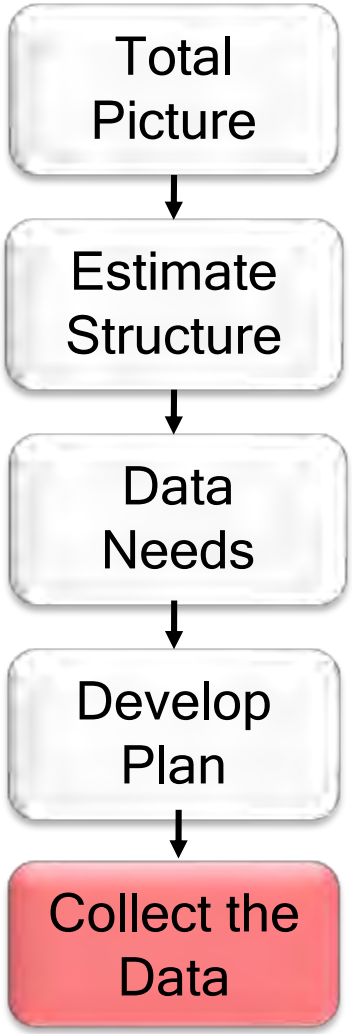
Data Collection Process – Develop Data Collection Plan



- Ensure sources identified for all needed cost elements
- Identify alternative/backup data sources
- Identify sources for cross-checks
- Establish a timeline/schedule
- Get plan approval

Tip: Only gather data that is appropriate to the task being estimated.

Data Collection Process – Collect the Data



- Conduct site visits at the source of the data.
- Utilize proven people skills to open doors
- Document everything in real time to capture key points
- Conduct Interviews with SMEs to provide proper context of the data.
- Question exactly what the data represents (Estimates, vs Actuals)
- Use your interviews as a lead to other key data holders and avenues of additional data.
- Verify data was collected as planned

Data Sources

- Requirements Documents (ICD, CDD, CPD)
- Baseline Descriptions (CARD, ICBD)
- Technical Databases (Satellite Crosslinks Database, etc.)
- Contractor-Provided Documentation (data sheets, etc.)
- Contractor Accounting System and Cost Reports (CPR, C/SSR, CFSR, C/SCSC, CCDR) 
- DACIMS for CCDRs and software metrics (SRDRs)
- DoD Historical Databases (VAMOSC, AFTOC, OSMIS, AF C3I H/W Maint Database, Navy OARS, JCARD, etc.)
- SARs and DAES for high-level schedule, technical and cost data, and particularly for changes to the program
- Labor Rate data from FPRAs, DCAA, DCMA

Data Sources (cont'd)

- Automated Cost Database (ACDB), an ACE-IT database for government users
- Other Organizations (FFRDCs, sister agencies)
- Other Information Systems (cost studies, GSA catalogs, documented cost estimates, etc.)
- Contracts and Cost Proposals
- President's Budget (PB)
- Testimony of Functional Specialists (SMEs)
- Competitive Intelligence, Market Analysis, Benchmarking
- The Internet! (Caveat Browser [*sic*])

Data Normalization

- Cost, Quantity and Sizing
- Content
- State of Development
- Key Groupings

11

20

How are you measuring?

What are you measuring?

When (and whom) are you measuring?

Which are you measuring?

"What's in the number?"
-Rick Collins [attributed]

Normalization – Cost, Quantity, and Sizing

- Cost Units
  - Corrected for price level changes associated with inflation (escalated)
  - Cost vs. price (and inclusion of burdens like OH or G&A)
  - Labor or material
  - $M vs. $K
  - Same year dollars
- Quantities
  - Unit or lot build
  - Flight, qualification, engineering model or spare
- Sizing Units
  - Weight
  - Density or volume
  - For software, physical vs. logical SLOC, DSLOC vs. ESLOC, counting methods, etc.

Cf. State of Development

Normalization Example: Escalation

Current Year to Base Year 2000 Dollars Conversion

5

You collect the following historical primary data:

Procurement cost for an AN/ARC-164 radio for a comparable aircraft shows actual costs to be $50,000 based on a May 1994 Bill of Materials.

Your estimate is being developed in Base Year 2000 dollars. Before you can use this data, it must be normalized.

The actual value you will want to use is $52,247 (BY2000$) for the radio.

Using the Air Force Procurement Appropriation index, what would $50K in 1994 CY$ be equivalent to in 2000 BY$?

The Answer:
$50,000 / 0.957 = $52,247 (BY2000$)

Normalization Example: Unit Conversion

Technical Sizing Data Conversion

9

You collect the following primary data:

Optical cable type 1: 5 Pounds Sterling (British) / meter
Optical cable type 2: $2.50 / linear foot

Since the tools you plan on using rely on $ per foot data, the above pieces of data need to be converted to $ per foot:

Normalized Result:

Cable type 1: $2.67 / foot

Cable type 2: $2.50 / foot

Assume 1 Pound Sterling = $1.75 (U.S.)
1 meter = 39.37 inches
1 foot = 12 inches

http://www.cnn.com/TECH/space/9909/30/mars.metric.02/

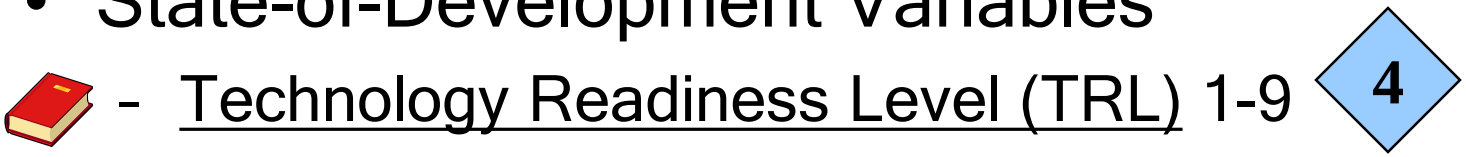
Warning: Even NASA has had an issue with unit conversion, resulting in the loss of a Mars Orbiter

Normalization - Content

- Cost Types
  - Recurring vs. Non-Recurring Costs
  - Fixed vs. Variable Costs
  - Helps account for:
    - Production quantity impacts
    - Time-phased impacts
- Adjust End Items for Homogeneity
  - Account for content differences in the end items
  - Missing or absent elements
  - Excess or inapplicable elements

Normalization – State of Development

- State-of-Development Variables
  - Technology Readiness Level (TRL) 1-9



  - Concept or Technology Demonstration Phase
  - Prototype and EMD
  - Beta I or II Test
  - Low Rate Initial Production (LRIP) or Full Rate Production (FRP)

    - Production units must be normalized to a consistent first unit (T1) or k^{th} unit

Note: Many of these terms are DoD distinctions but all development events have maturity stages related to production rates

Normalization – Key Groupings

- Products by Mission Application
  - Similar missions
  - Similar characteristics and traits
- By Operating Environment
  - Manned space
  - Unmanned space
  - Aerospace
  - Shipboard
  - Commercial systems

Aircraft Example: Normalization

- Cost Units
  - Data from different years inflated to common year
- Sizing Units
  - Data units are consistent; i.e., miles/hour vs. kilometers/hour
- End Items for Homogeneity
  - Tailor end items for jet engine – remove propeller costs
- State-Of-Development Variables
  - Was new product state-of-the-art, or just a small upgrade?
- Key Groupings
  - Classify according to jumbo jet, propeller, etc.
  - Note environment: turbine blades wear faster in desert sand
  - Costs incurred from development or production?

Data Normalization – Apples to Apples

	Question?	Apples to Apples	Normalization	Risks/Uncertainties/Perils
Cost	How are you measuring?	Apples last year, today, next year; Apples in U.S. or Canadian Dollars	Exchange rates, Escalation	Currency fluctuations , Commodity specific inflation indices
Quantity	How are you measuring?	One apple, a bag, a bushel or a truckload	Adjust for unit	Understatement of unit cost
Sizing	How are you measuring?	Pounds vs. kilograms	Unit conversion	No uncertainty, risk- or opportunity-like error
Content	What are you measuring?	Apple flesh, seeds, stems, skin	Addition or subtraction	Quantification of missing data
State of Development	When and whom are you measuring?	The seed, the sapling, the blossom, the apple, the applesauce	Adjustment factor	Step-down factors, CERs
Key Groupings	Which are you measuring?	Red apples vs. green, regular vs. organic, eating vs. baking	Indicator variables or segregation of data sets	Uncertainty stats from regression, possible increases in uncertainty due to smaller data set

Data Collection Summary

- Data are the lifeblood of cost estimating
- Different types of data for different purposes
- Data Collection follows a process
- Consistency, Comparability, Usability
- Where do you go for data?

What's in the number?

Resources – Books and Directives

- *AFSC Cost Estimating Handbook*, The Analytic Sciences Corporation, Reading, Massachusetts.
- *Parametric Estimating Handbook* (4th ed.), Joint Government/Industry Parametric Estimating Initiative (PEI), 2008
- AFSCP 173-4, Guide To Analysis Of Contractor Cost Data, September 1989
- DoD 5000.04-M-1 Cost and Software Data Reporting Manual, Nov. 4, 2011
  - http://www.dtic.mil/whs/directives/corres/pdf/500004m1.pdf
- "Cost Considerations in Systems Analysis," R-490-ASD, Gene H. Fisher, The Rand Corporation, December 1970.
- "An Estimation of USAF Aircraft Operating and Support Cost Relations," N-3062-ACQ, Gregory G. Hildebrandt, The Rand Corporation, May 1990.
- Military Equipment Cost Analysis, The Rand Corporation, 1971.
- *Metric Units and Conversion Charts: A Metrication Handbook for Engineers, Technologists, and Scientists*, Theodore Wildi, IEEE Press, 1995.
- RAND Military Acquisitions and Technology Publications
  - http://www.rand.org/publications/electronic/mat.html

Resources – Databases

- Air Force Total Ownership Cost (AFTOC)
  - https://acc.dau.mil/CommunityBrowser.aspx?id=404176
- Approved AF Factors
- Navy Visibility and Management of Operating and Support Costs (VAMOSC)
  - https://www.vamosc.navy.mil/
- Operating and Support Management Information System (OSMIS)
  - https://www.osmisweb.army.mil/
- Defense Cost and Resource Center (DCARC)
  - http://dcarc.cape.osd.mil/
- Joint Cost Analysis Research & Database (JCARD)
  - https://www.ncca.navy.mil/jcard/JCARD/login/login.cfm
- Open Architecture Retrieval System (OARS)
  - http://www.oars.navsea.navy.mil
- General Services Administration (GSA) catalog
  - http://gsaadvantage.gov

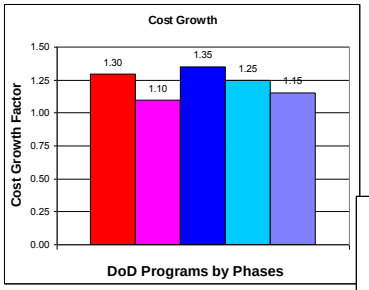
Related and Advanced Topics

- Uni, Bi, and Multi-variate Data
- NR/R Definition
- Technology Readiness Levels (TRLs)
- Data Collection for Risk Analysis

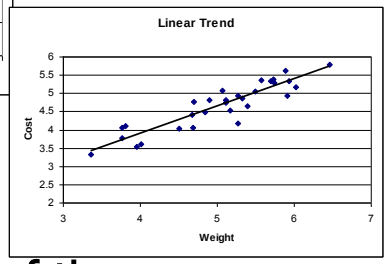
Data Types - Dimensionality

6

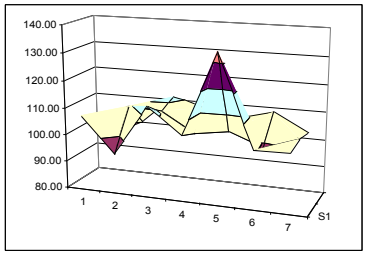
- Univariate
  - Cost data alone
    - Ex: A cup of coffee costs $1.00



- Bivariate
  - Cost with a single explanatory variable
    - Ex: Software development costs as a function of the number of lines of code



- Multivariate
  - Cost with multiple explanatory variables
    - Ex: Costs to operate a factory as a function of both fixed (overhead infrastructure and personnel) and multiple variable (material and direct labor) costs



Tip: Practically speaking, all data sets are Multivariate. These distinction are more apropos of data analysis than data collection.

NR/R Definition

- Non-Recurring Acquisition Costs:
  - Initial development
  - Setup costs
  - Upgrades and improvements
  - Applies to Hardware, Software, and Facilities
- Recurring Acquisition Costs:
  - Hardware production
  - Recapitalization
  - Installation of developed software
  - Spares
- DoD 5000.04-M-1 Cost and Software Data Reporting Manual gives a useful definition of both

Recurring costs are those subject to learning curve

7

TRLs 1-3 Descriptions

1. Basic principles observed and reported
   - Lowest level of technology readiness
   - Scientific research begins to be translated into applied research and development
   - Examples might include paper studies of a technology's basic properties
2. Technology concept and/or application formulated
   - Invention begins
     - Once basic principles are observed, practical applications can be invented.
   - Applications are speculative and there may be no proof or detailed analysis to support the assumptions
   - Examples are limited to analytic studies
3. Analytical and experimental critical function and/or characteristic proof of concept
   - Active research and development is initiated.
   - This includes analytical studies and laboratory studies to physically validate analytical predictions of separate elements of the technology
   - Examples include components that are not yet integrated or representative

TRLs 4-6 Descriptions

4. Component and/or breadboard validation in laboratory environment
   - Basic technological components are integrated to establish that they will work together
   - This is relatively "low laboratory environment fidelity" compared to the eventual system
   - Examples include integration of "ad hoc" hardware in the laboratory
5. Component and/or breadboard validation in relevant environment
   - Fidelity of breadboard technology increases significantly
   - The basic technological components are integrated with reasonably realistic supporting elements so they can be tested in a simulated environment
   - Examples include "high-fidelity" laboratory integration of components
6. System/subsystem model or prototype demonstration in a relevant environment
   - Representative model or prototype system, which is well beyond that of TRL 5, is tested in a relevant environment
   - Represents a major step up in a technology's demonstrated readiness
   - Examples include testing a prototype in a high-fidelity laboratory environment or in a simulated operational environment

TRLs 7-9 Descriptions

7. System prototype demonstration in an operational environment
   - Prototype near, or at, planned operational system
   - Represents a major step up from TRL 6, requiring demonstration of an actual system prototype in an operational environment such as an aircraft, vehicle, or space
   - Examples include testing the prototype in a test bed aircraft
8. Actual system completed and qualified through test and demonstration
   - Technology has been proven to work in its final form and under expected conditions
   - In almost all cases, this TRL represents the end of true system development
   - Examples include developmental test and evaluation of the system in its intended weapon system to determine if it meets design specifications
9. Actual system proven through successful mission operations
   - Actual application of the technology in its final form and under mission conditions, such as those encountered in operational test and evaluation
   - Examples include using the system under operational mission conditions

Data Collection for Risk Analysis

- Traditional cost estimating requires single snapshot of data
- Risk tries to address parameter growth
  - Need initial estimates and final actuals for:
    - Costs
    - Technical parameters
    - Schedule
  - Can examine trend of growth
- CERs tend to be final cost as a function of final weight (or other technical parameter)
  - Growth should be applied to technical parameter based on current state of design maturity