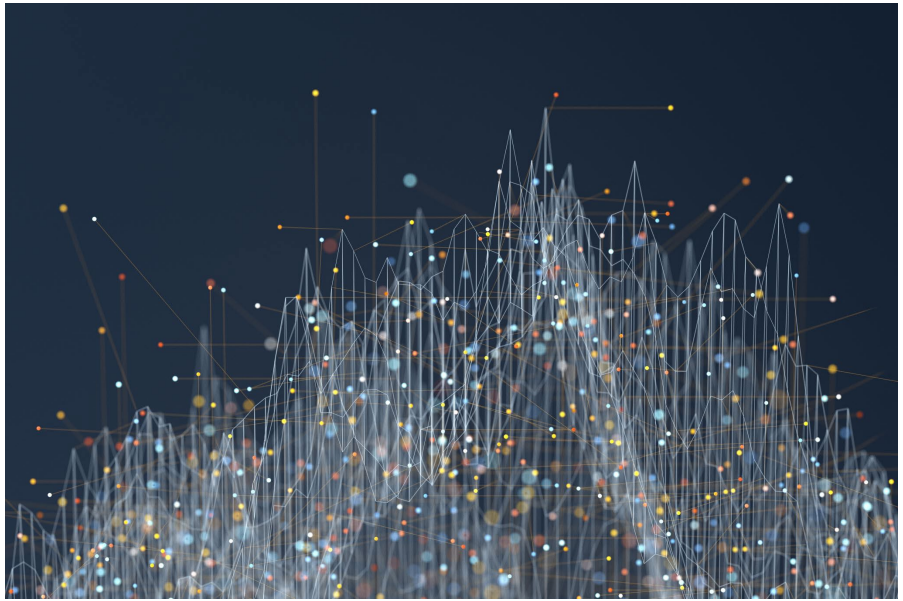
unison
Cost Engineering

Data Management for Cost Engineering Projects

Cara Cuiule

Agenda

- Introduction
- Data Types
- Storing Data
- Existing Data Repositories in Cost Estimation
- Case Study



Introduction

- Why store data in an organized repository?
- Data challenges
- Considerations for data management projects
  - Size of data
  - Type of data
  - Data transformation
  - Data security

Data Types

Data Types

- Structured Data
- Unstructured Data
- Semi-Structured Data

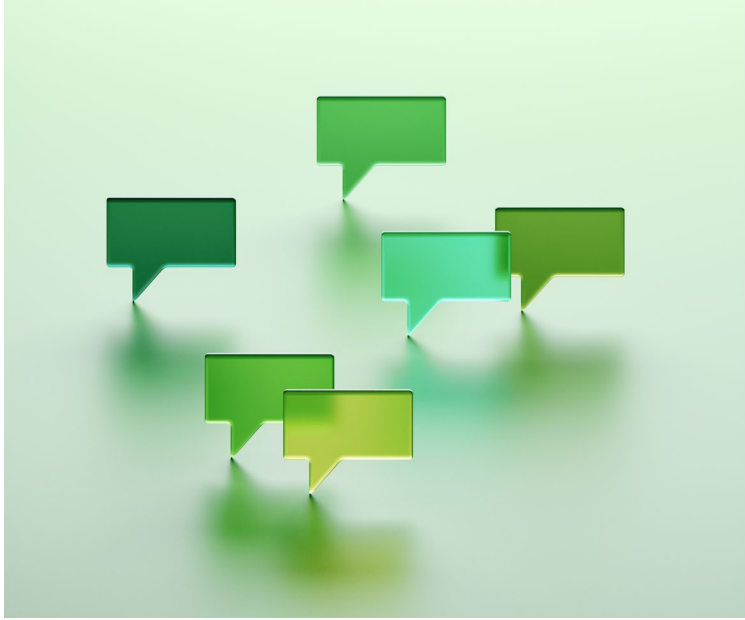
Structured Data

- Table = sets of rows and columns
- Fixed data types
- Tables can form Relational Databases
- **Best type of data for analysis**

Project ID	Phase	Start Date	End Date	System ID	Submission Date
Project_1	A	01/01/1999	01/01/2000	System_1	01/01/2010
Project_1	B	01/01/2000	01/01/2002	System_1	01/01/2010
Project_2	A	01/01/1999	01/01/2009	System_2	01/01/2010
Project_2	B	01/01/1999	01/01/2009	System_3	01/01/2010

Unstructured Data

- Unstructured data cannot be stored in a pre-defined column and row format
- Examples
  - Conference lecture recording
  - Open-ended survey question answer
  - A PDF of a contract document
- Unstructured data value: extracting structured data



Semi-Structured Data

- Data not in tables but has “tags” or “keys”
- Example: JSON demo File


```
[
    {
        "OrderOrLotID": "1",
        "CLIN_ID": "CLIN1",
        "EndItemID": "Var-A",
        "WBSElementID": "1.1.1",
        "AccountID": "FKWF-QPRF-FTNL",
        "NonrecurringOrRecurringID": "NONRECURRING",
        "FunctionalCategoryID": "DirEngLab1",
        "FunctionalOverheadCategoryID": "OverheadCategory1",
        "StandardCategoryID": "DIRECT_ENGINEERING_LABOR",
        "ReportingPeriodID": 1,
        "Tag1": "JVXRMC.SFMXND.WQFVMY",
        "Tag2": "WQCCFB.SQXSYG.SBNRNT",
        "Tag3": "FKCYJT.SQMDPH.YRPLZV",
        "Value_Dollars": 9519,
        "Value_Hours": 118
    },
    {
        "OrderOrLotID": "1",
        "CLIN_ID": "CLIN1",
        "EndItemID": "Var-A",
        "WBSElementID": "1.1.1",
        "AccountID": "FKWF-QPRF-FTNL",
        "NonrecurringOrRecurringID": "NONRECURRING",
        "FunctionalCategoryID": "DirEngLab1",
        "FunctionalOverheadCategoryID": "OverheadCategory1",
        "StandardCategoryID": "DIRECT_ENGINEERING_LABOR",
        "ReportingPeriodID": 2,
        "Tag1": "JVXRMC.SFMXND.WQFVMY",
        "Tag2": "WQCCFB.SQXSYG.SBNRNT",
        "Tag3": "FKCYJT.SQMDPH.YRPLZV",
        "Value_Dollars": 12730,
```


https://cade.osd.mil/tools/csdr-tools

Semi-Structured Data Transformation

- If semi-structured data has a lot of common tags, then a structured version will be very complete

```
{
    "OrderOrLotID": "1",
    "CLIN_ID": "CLIN1",
    "EndItemID": "Var-A",
    "WBSElementID": "1.1.1",
    "AccountID": "FKWF-QPRF-FTNL",
    "NonrecurringOrRecurringID": "NONRECURRING",
    "FunctionalCategoryID": "DirEngLab1",
    "FunctionalOverheadCategoryID": "OverheadCategory1",
    "StandardCategoryID": "DIRECT_ENGINEERING_LABOR",
    "ReportingPeriodID": 1,
```

A	B	C	D	E	F	G	H
OrderOrLot ID	CLIN_ID	EndItem ID	WBS Element ID	AccountID	Nonrecurring Or RecurringID	Functional CategoryID	FunctionalOverhead CategoryID
1	CLIN1	Var-A	1.1.1	FKWF-QPRF-FTNL	NONRECURRING	DirEngLab1	OverheadCategory1
1	CLIN1	Var-A	1.1.1	FKWF-QPRF-FTNL	NONRECURRING	DirEngLab1	OverheadCategory1
1	CLIN1	Var-A	1.1.1	FKWF-QPRF-FTNL	NONRECURRING	DirEngLab1	OverheadCategory1
1	CLIN1	Var-A	1.1.1	FKWF-QPRF-FTNL	NONRECURRING	DirEngLab1	OverheadCategory1
1	CLIN1	Var-A	1.1.1	FKWF-QPRF-FTNL	NONRECURRING	DirEngLab1	OverheadCategory1
1	CLIN1	Var-A	1.1.1	FKWF-QPRF-FTNL	NONRECURRING	DirEngLab1	OverheadCategory1
1	CLIN1	Var-A	1.1.1	FKWF-QPRF-FTNL	NONRECURRING	DirEngLab1	OverheadCategory1
1	CLIN1	Var-A	1.1.1	FKWF-QPRF-FTNL	NONRECURRING	DirEngLab2	OverheadCategory1
1	CLIN1	Var-A	1.1.1	FKWF-QPRF-FTNL	RECURRING	DirEngLab2	OverheadCategory1
1	CLIN1	Var-A	1.1.1	FKWF-QPRF-FTNL	RECURRING	DirEngLab2	OverheadCategory1
1	CLIN1	Var-A	1.1.1	FKWF-QPRF-FTNL	RECURRING	DirEngLab2	OverheadCategory1
1	CLIN1	Var-A	1.1.1	FKWF-QPRF-FTNL	RECURRING	DirEngLab2	OverheadCategory1
1	CLIN1	Var-A	1.1.1	FKWF-QPRF-FTNL	RECURRING	DirEngLab2	OverheadCategory1
1	CLIN1	Var-A	1.1.1	FKWF-QPRF-FTNL	RECURRING	DirEngLab2	OverheadCategory1
1	CLIN1	Var-A	1.1.1	FKWF-QPRF-FTNL	RECURRING	DirEngLab2	OverheadCategory1
1	CLIN1	Var-A	1.1.1	FKWF-QPRF-FTNL	RECURRING	DirEngLab2	OverheadCategory1
1	CLIN1	Var-A	1.1.1	FKWF-QPRF-FTNL	RECURRING	DirEngLab2	OverheadCategory1
1	CLIN1	Var-A	1.1.1	FKWF-QPRF-FTNL	RECURRING	DirEngLab2	OverheadCategory1

Storing Data

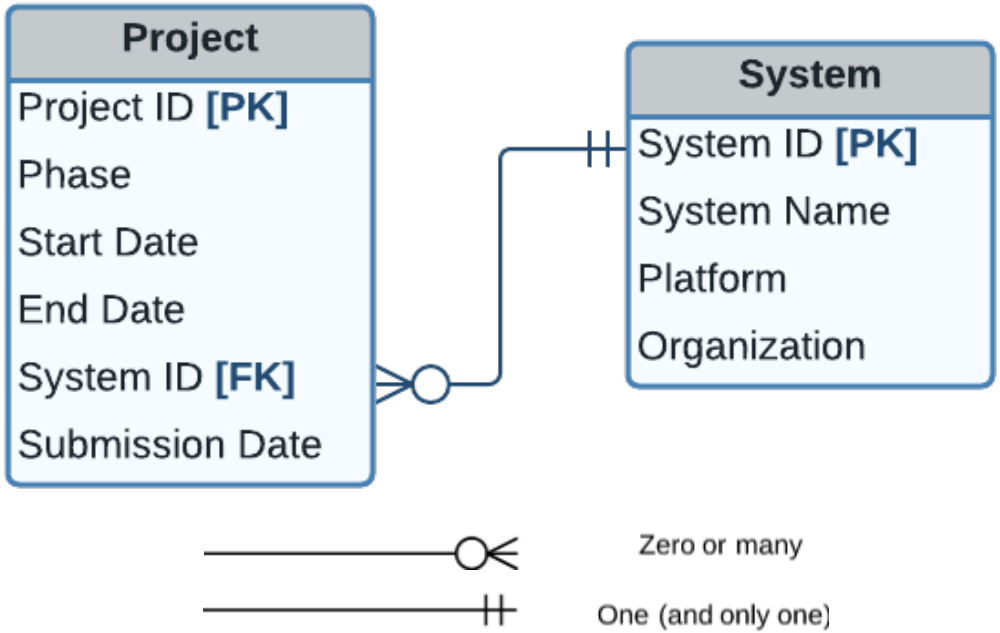
Relational Tables: Structured Data

Project ID	Phase	Start Date	End Date	System ID	Submission Date
Project_1	A	01/01/1999	01/01/2000	**System_1**	01/01/2010
Project_1	B	01/01/2000	01/01/2002	**System_1**	01/01/2010
Project_2	A	01/01/1999	01/01/2009	**System_2**	01/01/2010
Project_2	B	01/01/1999	01/01/2009	**System_3**	01/01/2010

System ID	System Name	Platform	Organization
Sys_1	Drone Variant X	Unmanned Air	Org A
Sys_2	Drone Variant Y	Unmanned Air	Org A
Sys_3	Rotorcraft Variant B	Manned Air	Org B

Entity Relational Diagram

- One-to-many relationship (most common)



Pros/Cons of Relational Databases

Pros

- Organize data
- Easily set data, security parameters in Relational Database Management System

Cons

- Can only store structured data
- Changing data structure or security after initialization is difficult
- Need to learn SQL to use it

Storing Unstructured and Semi-Structured Data

- Store in non-relational databases (NoSQL)
- Transforming into structured data takes extra effort/resources

Storing Large Amounts of Data

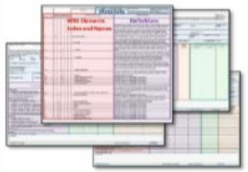
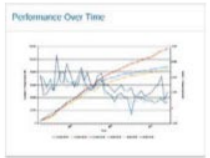
	Data Warehouse	Data Lake
Data Structure	Structured - collection of relational databases	Collection of any type of data
Schema Creation	Schema is created before data is written to the repository (“schema-on-write”)	Schema is created when the data is read from the repository (“schema-on-read”)
Structure Flexibility	Not Agile, takes more effort to add new sources of data	Very Agile, new sources of data can be added quickly
Type	SQL	NoSQL
Processing Speed	Faster	Slower
Normalization Method	Extract – Transform – Load (ETL)	Extract – Load – Transform (ELT)

Adapted From: https://towardsdatascience.com/what-is-a-hybrid-data-lake-b7ef2c3cce0c
https://www.dell.com/en-us/blog/schema-read-vs-schema-write-started/

Existing Data Repositories in Cost Estimation

- Federal Logistics Information System (FLIS)
- Cost Assessment Data Enterprise (CADE)

Cost	Earned Value (EVM)	Acquisition	Technical	Library
› Browse/Export Prime and Subcontractor Cost Data	› Browse/Export Data on ACAT I Prime Contracts › Quick-look Visualization Tools	› Program Information › SAR/MAR Annual Funding › SAR/MAR Schedule Events › CARDs*	› Software Database › Electronic CARDs (eCARDs)* › Technical Data Reports*	› Cost Estimates › Funding Memos › Program Briefings › Research Studies


https://www.dau.edu/Lists/Events/Attachments/47/08-09-2017%20DAU-Lunch-Learn-CADE-final_MTaylor.pdf

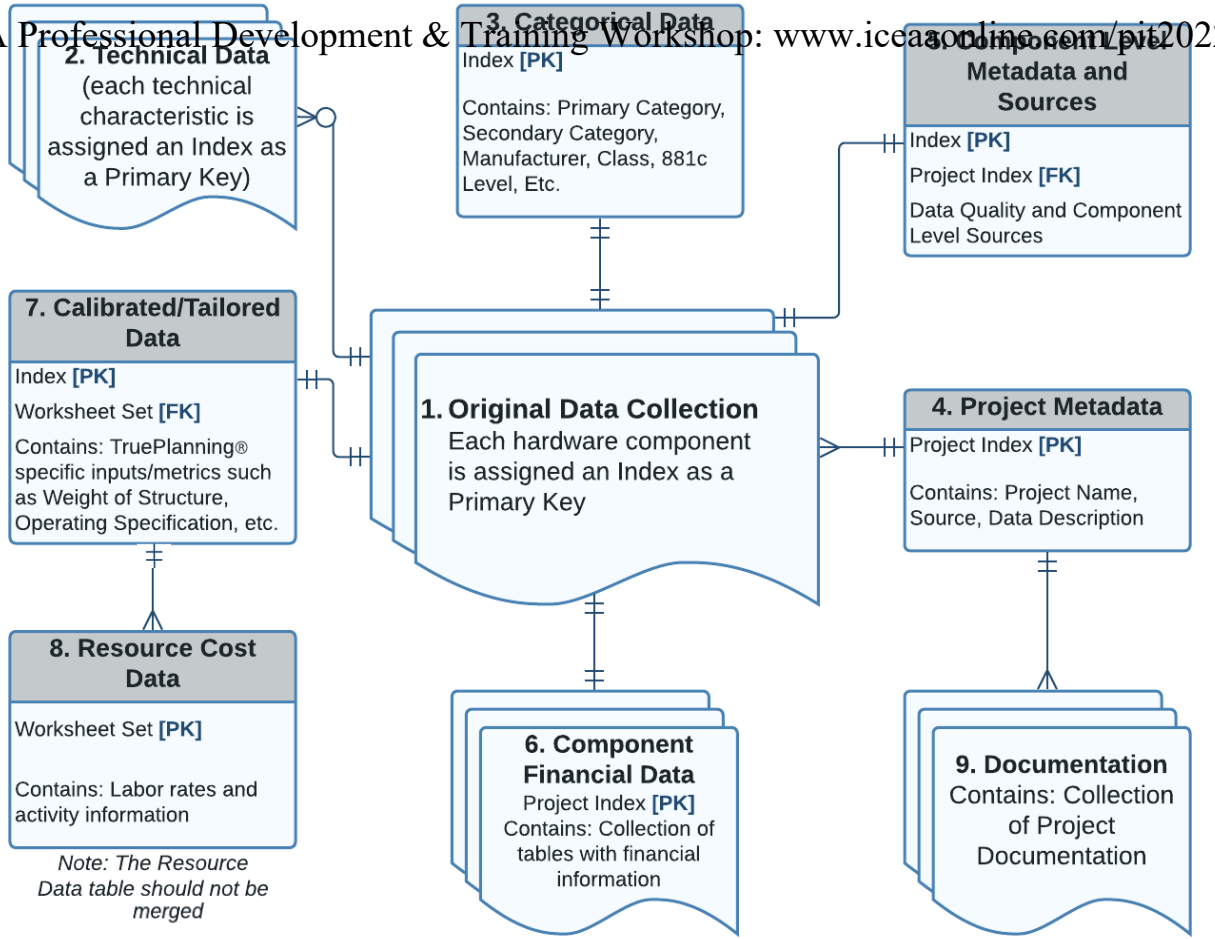
Case Study

Case Study

- Collection of data being gathered for hardware components
- Work in progress

Data Category	Data Type
1. Original Data Files	Any
2. Technical Data	Semi-Structured Collection of Tables
3. Categorical Data	Structured
4. Project Metadata	Structured
5. Component Level Data Sources and Data Quality	Structured
6. Component Financial Data	Structured
7. Calibrated/Tailored Data	Semi-Structured Collection of Tables
8. Resource Cost Data	Structured
9. Documentation	Any

Case Study Diagram



Collection of Documents with Unknown Structure

1. Original Data

9. Documentation

	A	B	C	D	E	F
1	Reference/ AN NUMBER	Issue	Source	Nato_Stock_ Number	Releasable	Type Functio
2	10M2L		RADCOM		YES	
3	10M3L	Original	RADCOM		YES	
4	10M4LDX	Original	RADCOM		YES	
5	124WB		Cushcraft catalogue		YES	
6	13B2		Cushcraft catalogue		YES	
7	15M4LDX	Original	RADCOM		YES	
8	17B2		Cushcraft catalogue		YES	
9	224WB		Cushcraft catalogue		YES	
10	26B2		Cushcraft catalogue		YES	
11	2M5L		RADCOM		YES	
12	2M7L		RADCOM		YES	
13	6M5DX	Original	RADCOM		YES	
14	6M5L	Original	RADCOM		YES	
15	6M7LDX	Original	RADCOM		YES	
16	6ML	Original	RADCOM		YES	

GSA Price List Effective 03/15/2018, Mod PS-0035

FLIR

GENERAL SERVICES ADMINISTRATION

Federal Supply Service
Authorized Federal Supply Schedule Price List

On-line access to contract ordering information, terms and conditions, up-to-date pricing, and the option to create an electronic delivery order are available through GSA*Advantage!*®, a menu-driven database system. The INTERNET address GSA *Advantage!*® is: GSAAdvantage.gov.

PROFESSIONAL AUDIO/VIDEO TELEMETRY/TRACKING, RECORDING/REPRODUCING AND SIGNAL DATA SOLUTIONS
FSC Group: 58I
Special Item Numbers: 58-6 and 58-7

CONTRACT NUMBER GS-03F-099CA

For more information on ordering from Federal Supply Schedules click on the FSS Schedules button at www.gsa.gov

FLIR SURVEILLANCE, INC.
GSA PRICE LIST
Contract No. GS-03F-099CA
PS-0035
Effective Date: MAR 15, 2018

FLIR

Part Number	Part Name	Description	GSA Price
TRNG 2LRU OL/USA	TRNG 2LRU/OPS/LM/CONUS/3DAY/8STUDENTS	TRNG 2LRU/OPS/LM/CONUS/3DAY/8STUDENTS Operator and Line (OL) Maintenance training (TRNG) for 2 LRU systems conducted within the Continental United States (CONUS). Operators & Line Maintenance course is a Level I certification course designed for Operators and Organizational level of maintainers who are charged with operation, system availability, and maintenance for FLIR's imaging systems. Operator training section is designed to provide operators and flight crews the ability to maximize performance and utilization. Curriculum allows operator to experience hands-on navigation through menus, enhance image optimization by applying proper settings under given operational limitations or environmental conditions. Maintenance section covers all organizational maintenance tasks including removal and replacement of LRUs, inspection and servicing requirements, plus testing and troubleshooting techniques. Course duration is 3 days for 8 students, at the customer location utilizing the customer owned gimbaled product. Certificate of Certification will be issued upon completion of the course and demonstrated knowledge of course materials through written and practical (hands-on) examinations. Course deliverables include an electronic Operator Manual and Student Training Guide. Course includes all instructor travel expenses.	$ 19,208
TRNG 2LRU OL/INTL	TRNG 2LRU/OPS/LM/OCONUS/3DAY/8STUDENTS	TRNG 2LRU/OPS/LM/OCONUS/3DAY/8STUDENTS Operator and Line (OL) Maintenance training (TRNG) for 2 LRU systems conducted Outside of the Continental United States (OCONUS). Operators & Line Maintenance course is a Level I certification course designed for Operators and Organizational level of maintainers who are charged with operation, system availability, and maintenance for FLIR's imaging systems. Operator training section is designed to provide operators and flight crews the ability to maximize performance and utilization. Curriculum allows operator to experience hands-on navigation through menus, enhance image optimization by applying proper settings under given operational limitations or environmental conditions. Maintenance section covers all organizational maintenance tasks including removal and replacement of LRUs, inspection and servicing requirements, plus testing and troubleshooting techniques. Course duration is 3 days for 8 students, at the customer location utilizing the customer owned gimbaled product. US State Dept approval / TAAs, as applicable, must be in place prior to instructor travel. Certificate of Certification will be issued upon completion of the course and demonstrated knowledge of course materials through written and practical (hands-on) examinations. Course deliverables include an	$ 26,470

Comprehensive Tables of Space Hardware Manufacturing Complexities

Purpose: Create manufacturing complexity references relevant to common space systems breakdown structures such as NASA WBS, USAF SMSC WBS, and Mil-Std 881c .

Background: A table of PRICE System manufacturing complexities was created from the following three data sources:

- PRICE KnowledgeNetwork (PRICE KN): A data source of specific product information provided to PRICE Estimating Suite (PES) customers from approximately 1995 through 2005. Since 2005, KN has been archived in MS Excel format. KN is over 12,200 records of individual hardware components and over 1,000 records of individual software records. Data covers all operating platforms; e.g. military ground, commercial, aircraft, space; over 3,500 of the hardware records deal with space components. Weights and manufacturing complexities are among the product information of each hardware record.
- Spacecraft Bus Component Calibrations: Selected Unmanned Space Cost Model (USCOM) data

Overview of Process

- Download Haystack data using "MCPLX Analysis" report
  - We'll review methods to find the right dataset.
- Load into Haystack Analyzer tool
  - Processes raw data file, applies a set of rules to cleanse, normalize, transform data
- Review the results, apply manual adjustments
  - The data is typically messy, and needs much effort to prep for calibration
  - Remove data that is clearly miscategorized or incorrect (i.e. circuit card weighing 10000 lbs, engine that costs $0.99, etc.)
  - Op Spec clues are gathered, but requires review
  - Quantities - examine procurement history, and decide on reasonable lot quantities for datapoints without procurement history available
- Run calibration
- Validate

TruePlanning by PRICE Systems 2

Collection of semi-structured data

- 2. Technical Data
  - Each data source has unique technical characteristics that cannot be anticipated ahead of time

Index	Caliber	Capacity
GunAmmo_120	.243 Win	10
GunAmmo_121	.308 Win	10
GunAmmo_122	.300 Blackout	5

Index	AVERAGE POWER RATING PER CHANNEL	BASIC SHAPE STYLE	ATTENUATION IN DECIBELS
000014072	27.0 WATTS OUTPUT		
000033553	1.0 KILOWATTS OUTPUT	1 RECTANGULAR OR SQUARE	
000033559	20.0 WATTS OUTPUT	RECTANGULAR OR SQUARE	4.0

- 6. Financial data
  - Can come in a variety of currencies and Fiscal Years, usually Unit Production Cost

Index	Unit Production Cost (Currency;£,GBR,2003;Metric)
MoD_103	99.95
MoD_104	169.95
MoD_107	799
MoD_109	536.6220705

Index	Unit Production Cost (Currency;$,USA,1984;Metric)
MoD_100	75.13627277
MoD_101	105.1907819
MoD_102	145.2634607

3. Categorical Data

- Structured version of technical data
- Includes (but not limited to);
  - Categorization
  - 881c Level
  - Proprietary

Index	Primary Category	Class	881c Level1	Proprietary
KN_1	Airframe	Military	Aircraft System	No
KN_2	Fuselage	Military	Aircraft System	No
KN_3	Window Assembly	Military	Aircraft System	No
KN_4	Spoiler	Military	Aircraft System	No
KN_5	Aileron	Military	Aircraft System	No
KN_6	Throttle Quadrant	Military	Aircraft System	No
KN_7	Wing Flaps	Military	Aircraft System	No

4. Project Metadata

- Structured version of technical data
- Includes (but not limited to);
  - Project Index
  - Project Name
  - Data Description
  - Year of Study
  - TP Calibration Version

Project Index	Project Name	Project Data Description	Year of Study	TruePlanning Calibration Version
GunAmmoProj	Gun and Ammunition Item Lists	Firearms and Ammunition data webscraped from Internet	2018	16.2
KN	TruePlanning Knowledge Network	Legacy open source data from PRICE Systems.	Not Available	Not Available
EO_IR	FLIR Item List	FLIR data gathered from the internet	2018	16.2

5. Component Level Metadata

- Structured version of technical data
- Includes (but not limited to);
  - Component Index
  - Project Index
  - Source
  - Data Quality
  - Date

Index	Project Index	Source	Data Quality - Total Weight	Date
GunAmmo_1002	GunAmmoProj	https://www.hyattgunstore.com/heckler-koch-vp9-flat-dark-earth-9mm-pistol-with-standard-sights.html	Green	4/12/2018
GunAmmo_1003	GunAmmoProj	https://www.hyattgunstore.com/heckler-koch-vp9-flat-dark-earth-slide-od-green-frame-9mm-pistol-with-standard-sights.html	Green	4/12/2018
GunAmmo_1004	GunAmmoProj	https://www.hyattgunstore.com/heckler-koch-vp9-9mm-pistol-with-standard-sights.html	Green	4/12/2018
GunAmmo_1005	GunAmmoProj	https://www.hyattgunstore.com/heckler-koch-vp9-sk-subcompact-9mm-pistol-with-night-sights-and-three-magazines.html	Green	4/12/2018

Model Specific Data

- 7. Calibrated/Tailored Data
  - Over 200 Columns

Index	Worksheet Set	Unit Production Cost (Currency;$,USA, 2018;Metric)	Total Weight (Weight;lbs; Metric)	Weight of Structure(Weight; lbs; Metric)	Weight of Electronics (Weight; lbs; Metric)
GunAmmo_2	GunAmmoWS	1575.956	6	6	0
GunAmmo_3	GunAmmoWS	1566.889	5.15	5.15	0

- 8. Resource Cost Data

Worksheet Set	Cost Object	Activity	Resource	Country	Unit Cost	Cost Unit
GunAmmoWS	Hardware Component	Development Engineering		United States		
GunAmmoWS	Hardware Component	Development Engineering	Design Engineering	United States	98088.34	$/Year
GunAmmoWS	Hardware Component	Development Engineering	System Engineering	United States	127403.67	$/Year

Conclusion

Conclusion

- Data management for cost estimation organizes data based on differences
  - Data format
    - Structured, unstructured, semi-structured
  - Data collected
    - Technical, Cost, Calibrated, Metadata, etc.
- Choose tools based on project/data requirements **before** the project is started
- Might need additional resources (IT Department) to implement