4° International Conference on
IT Data collection, Analysis and Benchmarking
Los Angeles, CA (USA) – September 7, 2016

Data Driven Cost Estimating

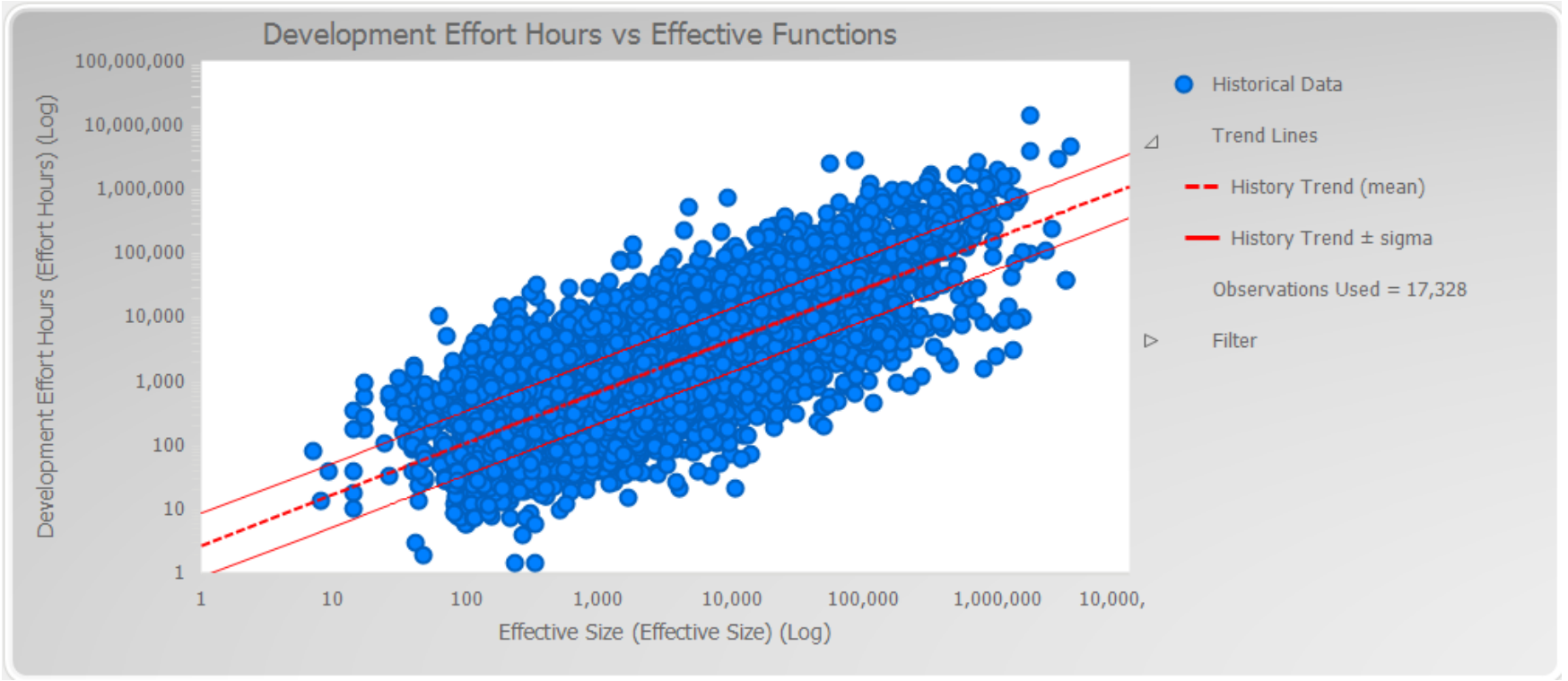
And the Role of Industry and Private Data

Karen McRitchie

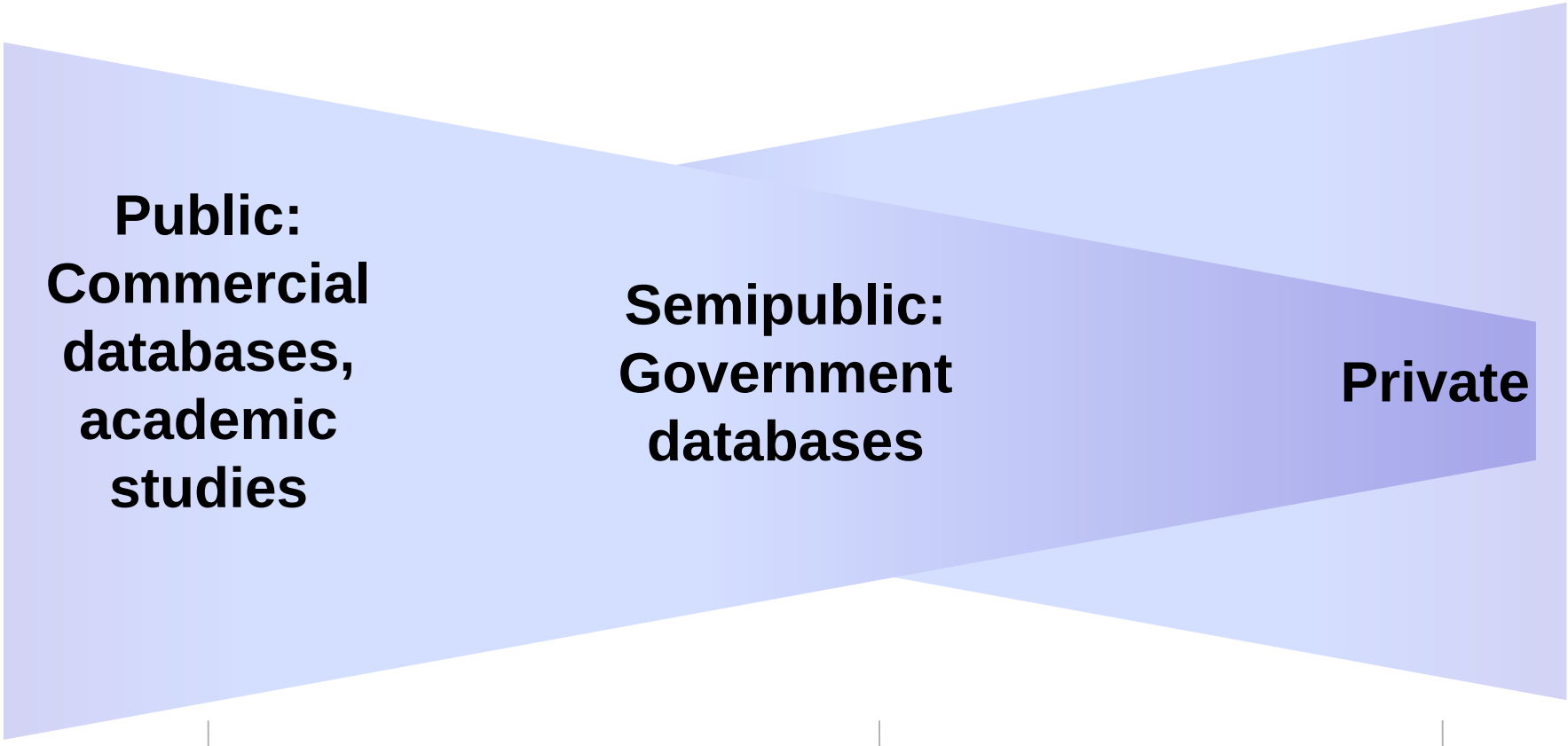
Galorath Incorporated

In the Beginning

Now


Development Effort Hours vs Effective Functions
Development Effort Hours (Effort Hours) (Log)
100,000,000
10,000,000
1,000,000
100,000
10,000
1,000
100
10
1
1
10
100
1,000
10,000
100,000
1,000,000
10,000,
Effective Size (Effective Size) (Log)
Historical Data
Trend Lines
History Trend (mean)
History Trend ± sigma
Observations Used = 17,328
Filter

Sources of Data



Public versus Private – Comparisons Between…

Public Data	Private Data
Can be **shared**	Must be **sanitized**, if shareable at all
Supposedly **diverse**	Decidedly **not diverse**
Typically **"narrow"** data coverage	Typically **"wider"** data coverage
Project knowledge usually general	Project knowledge can be **intimate**
Several **Categorical** labels may fit	**Categorical** labels are more specific

Data Gathering – Beggars Can't Be Choosy

A List of "Nice to Haves"

Scope generally is considered most important; all else is negotiable. Though even scope can be traded away…

- Platform
- Application
- Scope
- Origin
- Effort
- Duration
- Cost
- Included labor
- Included activities
- Defects
- Staffing
- Complexity
- Extraordinary circumstances

Private Data – Bonus
Data Collection Over Time

1. **Up front:**
   To gauge scope creep from start to finish. Very important for calibration, since project estimates also are made <u>at the beginning.</u>
2. **In process:**
   During development for management, for internal use in identifying issues and gauging progress
3. **Post mortem:**
   Upon completing development, this becomes the bedrock for a project repository
4. **In service:**
   During maintenance to gauge life- cycle costs

Private – The "Data Under Duress" Scenario

"The Data…"

- **"…may air our dirty laundry."**
- **"…reveals our labor rates."**
- **"… is not relevant or is different from other projects"**
- **"…doesn't exist."**
- **"(…doesn't exist, although we can't tell you that.)."**
- **"…is not clean."**
- **"…may threaten our bid strategy."**

Data Sharing – Sunlight Is the Best Policy

- **Your laundry isn't so dirty. Everyone knows developing software is hard and takes more time than planned.**
- **You don't need to reveal labor rates.**
- **Your competitors won't ever see your data.**
- **You have more data than you know, even if you don't record effort and scope.**
- **Let us save you time and try and clean it for you.**
- **Resultant data products help you bid more successfully.**

Data Aggregating – A Game of Least Common Denominator

Combining data, you rely on whatever all the data sets share.

➢ Platform	➢ Platform	➢ Platform	→ Platform
	➢ Application	➢ Application	
➢ Scope	➢ Scope	➢ Scope	→ Scope
➢ Origin			
➢ Effort	➢ Effort	➢ Effort	→ Effort
➢ Duration		➢ Duration	
	➢ Cost	➢ Cost	
➢ Included labor	➢ Included labor		
		➢ Included activities	
➢ Defects	➢ Defects		
	➢ Staffing		
➢ Complexity	➢ Complexity	➢ Complexity	→ Complexity
➢ Circumstances	➢ Circumstances	➢ Circumstances	→ Circumstances

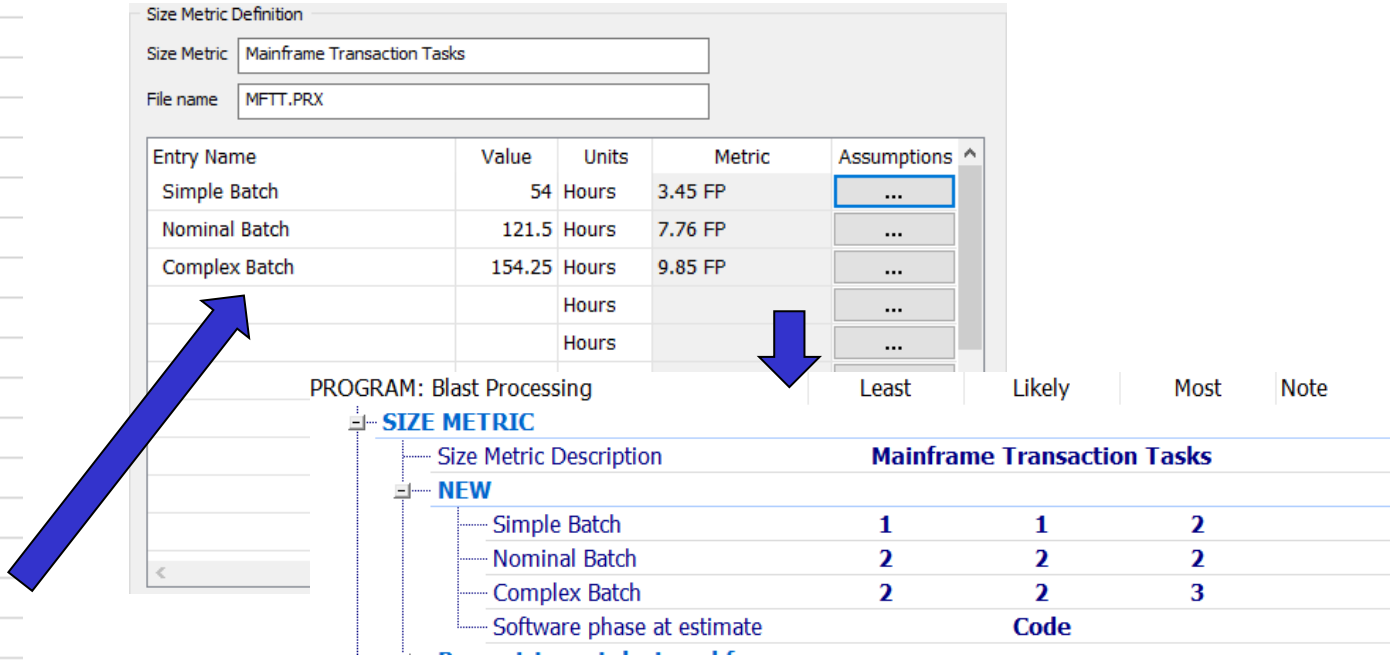
Although with opportunity for side studies.

Task/Artifact Based Sizing - Creating Custom Metrics for Estimation

Create custom metrics by 'back firing' from accounting data. We support this offline and now directly in SEER-SEM.

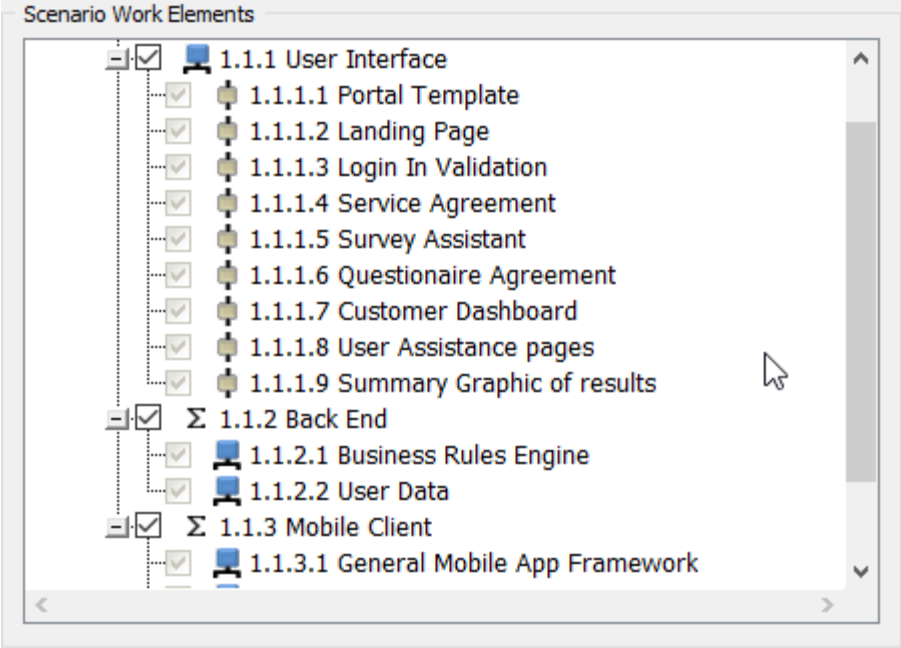
DATA	
Task	Hours
PX1-1758	133
PX1-3658	28
PX0-7883	57
PX1-1913	115
PX1-6462	112
PX0-4531	53
PX0-3619	210
PX1-7922	16
PX0-3973	126
PX0-4620	8
PX0-3311	192
PX0-2488	117
PX1-4917	130
PX1-3708	140
PX0-4318	159
PX0-5449	204

SUMMARY METRICS	
Max	210
75%	154.25
50%	121.5
25%	54
Min	8



Meta-Data

- Capture Project Patterns for estimation
  - Consistent structures make comparing projects easier
  - Avoids errors of omission in estimation
  - Facilitates data collection by identifying common categories



Dynamic Calibration – making adjustments on the fly

Project Calibration Data

Current Database Demographics

Program: Process Simulation

Platform	Business and Non-Critical
Application	Simulation
Total Points In Selected Database	7,840
Points Matching Platform and/or Application	3,425
Points Matching Platform Only	3,421
Points Matching Application Only	4
Points Matching Platform and Application	0

Calibration Selections

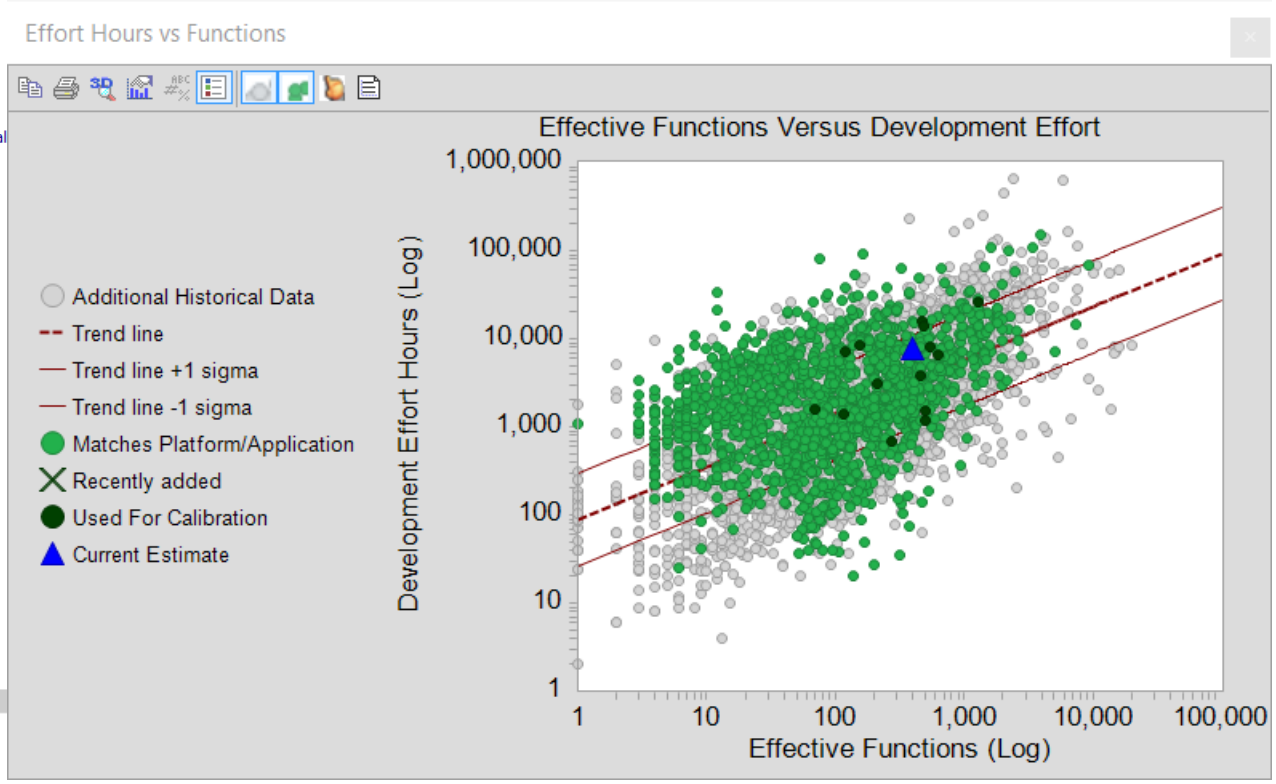
Calibration Points From Class Kbase	4 orig 0 removed
Points Manually Selected	10
Calibration Points In Current Database	14

Calibrations

Effort	0.99
Schedule	0.78

Points Used for Calibration and Available For Review

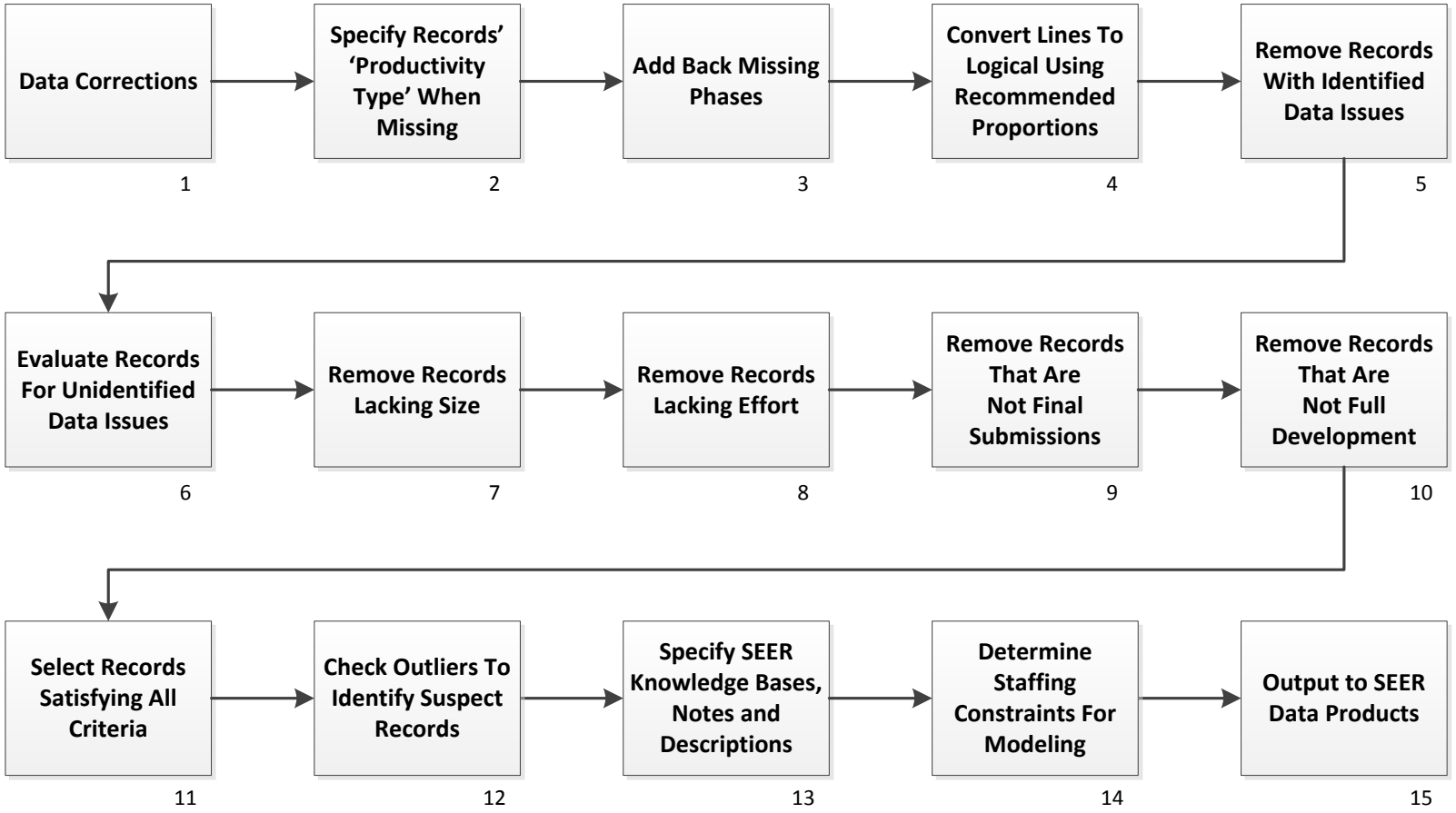
Name	Effort Hours
ISBSG D&E 10276	3,876
ISBSG D&E 10537	1,197
ISBSG D&E 10559	1,414
ISBSG D&E 10570	3,034
ISBSG D&E 10700	1,607



Static Calibration/CER Generation

- Requires data prep and data processing
- Running the regressions is the easy part
- Be prepared to rinse and repeat

Process Shared Data For Analysis



Forensics

Schedule. Was it "stop and start"? Were there schedule constraints?

Resources. Were there hard-hitting resource constraints?

Volatility. Did requirements undergo extraordinary evolution?

Manager's Objectives. Was it to complete the project in *minimum time* or at *least cost*?

Effort. Are effort figures actually derived from cost figures?

When creating the estimate --- adjustments for *extraordinary* conditions may be possible *within the software estimating model*.

Exploratory Data Analysis (ANOVA or regressions)

Helps to determine highest explanatories and functional forms.

VARIABLE NAME	ESTIMATED COEFFICIENT	T-RATIO	P-VALUE	DEFINITION OF VARIABLE
LR2	0.2842	5.52	0	1 when resource level is 2 or above * log UFPs
R2_DUM	-1.7423	-6.25	0	1 when resource level is 2 or above
LR4	0.0436	2.15	0.032	1 when resource level is 4 * log UFPs
FOURGL	-0.4255	-8.72	0	1 when 4GL language is used
LNUFP	0.6960	29.36	0	log of UFPs
PLAN	0.6842	2.38	0.017	1 when the Plan phase is included, according to revised list of phases
BUILD	-1.3634	-3.31	0.001	1 when the Build phase is included, according to revised list of phases
TST	1.1291	3.26	0.001	1 when the Test phase is included, according to revised list of phases
IMPL2	0.0907	1.68	0.094	1 when the Implementation phase is included, according to ORIGINAL list
LPLAN	-0.1511	-2.79	0.005	PLAN * log of UFPs
LSPEC	0.1249	7.33	0	SPEC * log of UFPs
LBUILD	0.1721	2.10	0.036	BUILD * log of UFPs
LTEST	-0.2344	-3.29	0.001	LTEST * log of UFPs
BUSINESS	-0.2778	-5.87	0	1 for non-critical Business projects
CLISERV	-0.1593	-1.76	0.078	1 for non-critical Client-Server projects
CONSTANT	4.0837	33.85	0	

Principal Components Analysis also is useful for reducing the number of variables.

Typical forms:

$$Effort = a * size^{entropy}$$

$$Duration = b * size^{duration\ entropy}$$

Classification Automation

Automated classification:

- Enables consistency
- Reduces mistakes

	BusiMissCrit	business	Client	clientsr
BusiMissCrit	1948	209	189	558
business		2408	257	720
Client			706	303
clientsr				1755

SECTOR / ORG / APPLICATION LOOKUP TABLE

Industry Sector	Organisation Type	Application Group	Application Type	co	BusiMissCrit	business	Client	clientsr	cloud	embedded	erp	finapplic	grounds
	All industry organization typ	Business Application	Security/Authentication;	All	1								1
	All-purpose;	Business Application	Document management;	All-purpose;		1							
	All-purpose;	Business Application	Document management;Job, case,	All-purpose;		1							
	All-purpose;	Real-Time Application	Telecom & network management;	All-purpose;Real-Time ApplicationTelecom & network management;									1
Banking	Agriculture, Forestry, Fishin	Business Application	Management Information System;	BankingAgri	1								
Banking	Banking;			BankingBanking;									
Banking	Banking;	Business Application		BankingBan	1								
Banking	Banking;	Business Application	Application Security Control;	Ba	1								
Banking	Banking;	Business Application	Auditing Management;	BankingBan	1								
Banking	Banking;	Business Application	Cards and Payments;	Ba	1							1	
Banking	Banking;	Business Application	Catalogue/register of things or ever	BankingBan	1								
Banking	Banking;	Business Application	Client Server;	BankingBanking;Business Applic			1						
Banking	Banking;	Business Application	Client/Server Customer Service apr	BankingBanking;Business Applic			1						
Banking	Banking;	Business Application	Customer billing/relationship manag	BankingBan	1								

ISBSG Corporate Release February 2015 mappings r14.xlsm

	A	B	C	D	E	F	G	IT	IU	IV	IW	IX	IY	IZ	JA	JB	JC	JD	JE	JF	JG	JH	JI	JJ	JK	JL	JM	JN	JO
1	ISBSG							PLATFORM DERIVATION																					
2	Release	6760 rows														Source & Rating	1		Source Rating:	1	2	3	4	5	6	Source & Rating	1		
3		Rating	Rating	Software Age	Major Grouping	Major Grouping	Major Grouping	BUSINESS AREA TYPE			Source & Rating	1				ARCHITECTURE			WEB DEVELOPMENT		DEVELOPMENT PLATF			Source & Rating	1	TARGET PLATFORM			
4	ISBSG Project ID	Data Quality Rat	UFP rating	Year of Project	Industry Sector	Organisation Type	Application Group																						
5	10001	D	A	1998	Service Indu	Recreation & Personnel Services;	Business Application	1	0	0	0	0	0	0	0				0			0	0	1	0				
6	10011	B	A	1996	Construction	Construction;	Business Application	0	0	0	0	0	0	0	0	1	0	0	0			0	0	0	1				
7	10012	B	A	2002	Wholesale & F	Billing;	Business Application												0										
8	10014	B	A	2004												1	0	0	0										
9	10015	B	A	2000	Wholesale & F	Wholesale & Retail Trade;	Business Application									1	0	0	0			0	1	0	0				
10	10026	B	A	2000	Insurance	Insurance;	Business Application	0	0	0	0	0	0	0	0				0			0	1	0	0				
11	10029	B		2004	Banking	Banking;	Business Application												0			0	1	0	0				

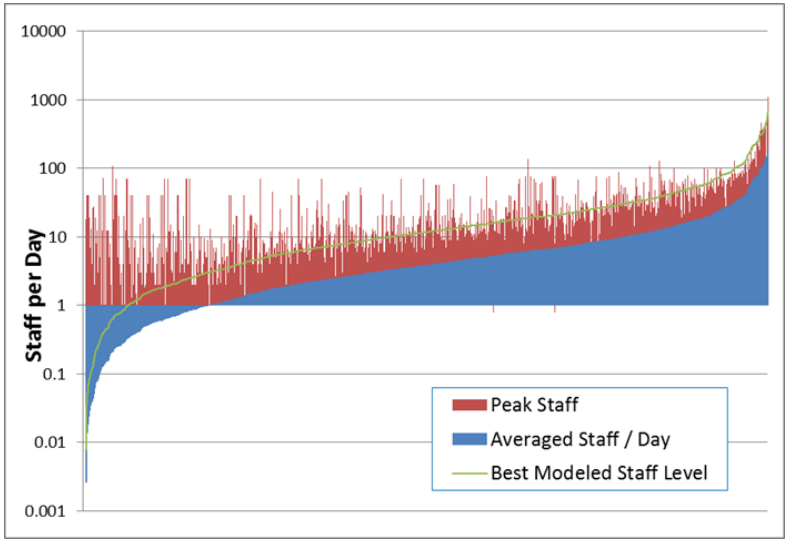
Typical Normalizations

- Adjustment to Line of Code or Function Point measures
  - Methods vary
  - For private data, frequently no adjustment is made – if customer prefers a given metric "flavor", we use it
- Adjustments to re-include missing activities or labor
  - Typically using simple proportions derived from other projects in the sample
- Inferring peak staffing
  - Informed by average staffing, when known or as calculated (example on next page)

Peak Staff - Significant, Often Missing

Staffing constraints have a significant effect on estimating; when not specified, they should be inferred when possible:

- Use reported peak staff when available
- When not reported, based on analysis, specify peak staff at X * average staff with no constraints below Y FTEs

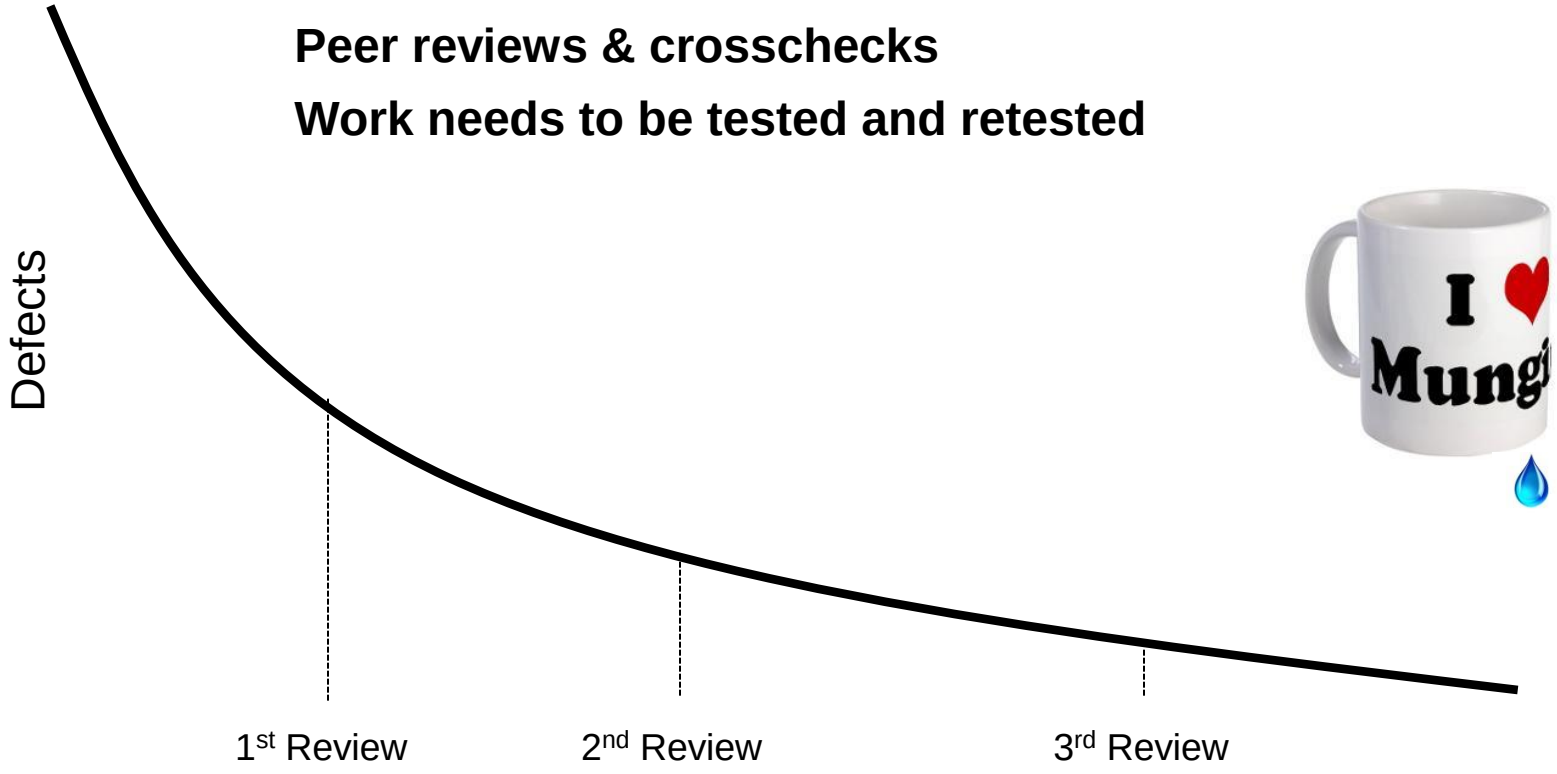


Oversight

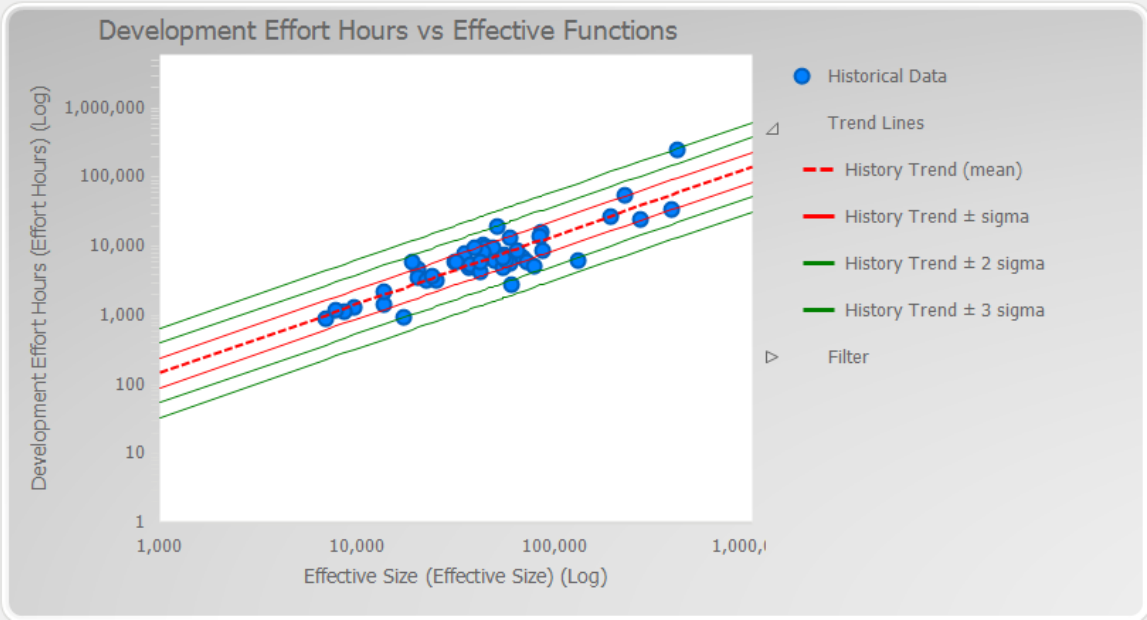
Data processing is like coding:

Peer reviews & crosschecks

Work needs to be tested and retested



Sanity Checks



SEER Metrics is used to plot data within probability bands.
Points outside 2 sigma are examined.

More vs. Less Data – Double-Edged Sword

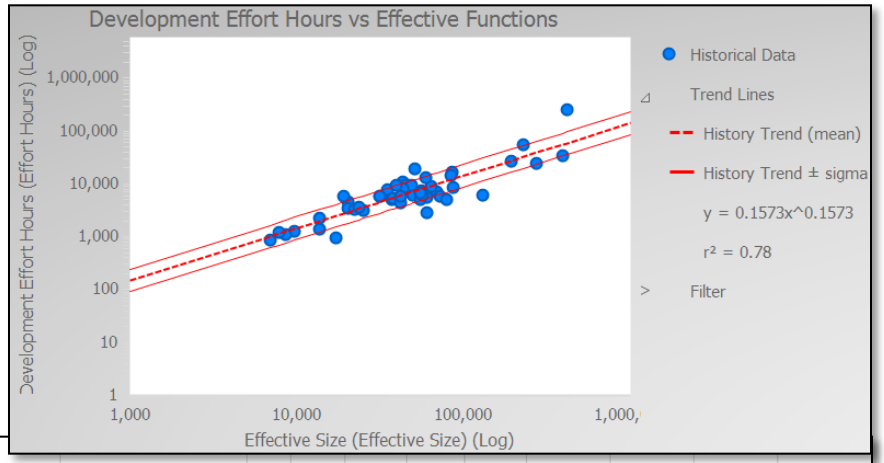
Less Data	More Data
Harder to spot outliers	**Easier** to spot outliers
Easier to explain them	**Harder** to explain them
Harder to systemically correct	**Easier** to systemically correct

Calibration Automation

Automation for solving form y=a*s^b

OLS using SEER Metrics

LAE using Excel



Customer Data Set

	Sample size	168

OLS regression results for comparison

Exponents	
Effort	Duration
1.0792	0.5243
Multipliers	
Effort	Duration
0.2895	0.0393

0 - A SEM project needs to be created beforehand

1 - Reset sheet and receive new data

Reset Sheet and Receive Data

2 - Obtain calibration values, interacting with SEM

Obtain Calibration Values

Effort Exponent	1.07
Schedule Exponent	0.53
Effort Multiplier	0.76
Schedule Multiplier	1.78

If you receive a message requesting a rerun, press this button

Re-Solve Multipliers

LAE REGRESSION

	Effort		Duration	
	LAE regression parameters		LAE regression parameters	
	multiplier	exponent	multiplier	exponent
	0.340463098	1.069530044	0.189147742	0.534121602
	min	2.6917	min	0.8547
	sum of errors	56.2768	sum of errors	122.2087
	Multipliers from SEM with LAE calibrations			
	0.4459		0.1061	
	Ratios to re-obtain LAE multipliers			
	0.763541		1.782731	

Effective Size	Development Actual Effort	Development Actual Schedule	Development Hours	Development Schedule Multiplier	Development Actual Schedule	Development Hours	Development Schedule Multiplier	COMBINED PERFORMANCE				log Effort	log Schedule	Log estimate	difference (logged)	Log estimate	difference (logged)
25900	17875.2	3.24	54109.81	32.84	3.24	39020.3	45.54	SEM Calibration Performance		Effort	Duration	4.2523	0.51022	4.2522	0.0000	1.6340	1.1238
24600	17875.2	31.60	50867.27	32.17	31.60	36928.39	4.43E+01	SEM uncalibrated	Median absolute deviation	min	122%	92%	4.2523	1.49973	4.2283	0.0239	1.6221	0.1224
7700	4742.4	4.08	12621.33	20.21	4.08	10656.26	2.39E+01	SEM uncalibrated	Mean absolute deviation	min	373%	1700%	3.676	0.61067	3.6888	0.0128	1.3527	0.7420

Measuring Calibration Efficacy

Calibration efficacy can be:

- **Specific** – "Does this stratification produce enough predictions within X percent?"
- **Relative** – "Does another stratification work better while capturing the requisite estimating scenarios?"

Stratification A				Stratification B			
		Effort	Duration			Effort	Duration*
Median absolute deviation	min	51%	78%	Median absolute deviation	min	37%	96%
Mean absolute deviation	min	80%	163%	Mean absolute deviation	min	48%	146%
Standard error (normalized)	min	88%	75%	Standard error (normalized)	min	90%	53%
Prediction within 10%	max	6%	13%	Prediction within 10%	max	18%	5%
Prediction within 25%	max	18%	25%	Prediction within 25%	max	34%	22%
R-squared	max	66%	14%	R-squared	max	94%	0%
Number of records:		109		Number of records:		103	
		Effort	Duration			Effort	Duration
Multiplier		0.95	2.04	Multiplier		0.90	5.04
Exponent		0.84	0.23	Exponent		0.93	0.10
Original multiplier		4.65	2.19	Original multiplier		2.06	23.20
OLS Multiplier		4.66	1.51	OLS Multiplier		3.65	23.20
OLS Exponent		0.83	0.26	OLS Exponent		0.89	0.05

Typical Statistics

Mean Magnitude of Relative Error (MMRE)	Percentage variation between actual and estimate with actual as divisor. *Closer to zero the better.*
Average Estimate Ratio	The average of (actual / estimate) ratios. *Closer to one the better.*
Median Estimate Ratio	The median of (actual / estimate) ratios. *Closer to one the better.*
Standard Deviation	The standard deviation of estimate vs. actual variances. *Closer to zero the better.*
Prediction (x)	The percent of estimates falling within *x%* of actual outcomes. *Closer to 100% the better.*
R-squared	Goodness of fit. (The goal is to purely "fit the data" and so this is a good measure.) *Closer to one the better.*
Other statistics	Chow for structural breaks (such as may occur as scope grows), F test to detect difference between two samples (such as estimate versus actual),

Side Studies

An example of the studies possible, in addition to classic size-effort and duration-effort relationships:

1. Size growth studies, to understand estimated versus actual system size
2. The impact of staffing on project productivity
3. Determinants of project duration
4. Ability of project requirements to determine effort and duration
5. Variations in productivity by program, division/contractor, maturity rating, programming language, toolset, etc.
6. Determinants of project phases' relative durations
7. Standardized, multiple tags for projects, yielding more descriptive information and more data for specific queries
8. Ability to predict maintenance and sustainment
9. What taxonomy has the best explanatory versus descriptive power?

Private/Local Data

- Use to shape the estimation prociess
  - Sizing
  - Project Patterns
  - Productivity Tuning
  - Economic Factors
  - Estimate Catalogs
- Focus is consistency, repeatibility and accuracy

Public/Industry Data

- Use to target major trends
  - Calibration/CER Generation
  - Productivity Benchmarking
- Sanity Check Your Estimate
- Vendor/Competitor Evaluation
- Evaluate Industry Demographics

Applying Data To Estimation

- It's not a lights out process
  - Requires analysts, SMEs, humans to process evaluate and identify where it can add value
- Getting started can be slow and bumpy
  - As the process is repeated, it gets easier
  - Automate as much as possible
- Revisit processes
- Meausre the benefits
  - Accuracy improvements, estimate turnaround, confidence in estimates