



***PRICE***<sup>®</sup>

*Webinar*

---

## Creating Knowledge from Data:

The PRICE<sup>®</sup> Data Initiative

Estimate with Confidence™

© 2021 PRICE Systems, L.L.C. All Rights Reserved

# PRICE®

## Cost Analytics



TrueExplorer



TrueFindings



PRICE® Models



TruePlanner



TrueMapper



TrueBOE



TrueXLS

Search & extract data from the PCA Ecosystem

Manage & Analyze Data Sets

Predictive Models

Integration Framework

Customer Data Mapping

Basis-of-Estimate Generator

Access PCA Engine from Excel



# Today's Presenter

## Arlene Minkiewicz

Chief Scientist

- 36 years of cost research experience
- Leads PRICE<sup>®</sup> Research, responsible for data research and analysis of costing trends.
- Designed and led the implementation efforts of the PRICE<sup>®</sup> Models, predictive analytic cost models.
- Works with industry leaders to collect and maintain cost research data, sharing analyses with the cost estimating community
- Her research has been published in books, trade journals and conference proceedings and has been awarded frequently by the industry. She speaks frequently at conferences on subjects of cost estimating and measurement.



# Today's Presenter



## Cara Cuiule

Cost Research Analyst

- Data Initiative Lead
- Proficiency in software cost estimation and automated data collection
- Instrumental in cost research projects, predictive model assessment, and customer project support
- Annual speaker at the ICEAA Workshop

# Overview

Background

Project Goals

Progress

- Data Collection & ETL
- Calibration Details
- Expert Judgement Elicitation

Live Demo

Results and Next Steps

Summary



# Background

- Customers want models to provide clear “in your face” feedback regarding whether a user input falls inside or outside of the PRICE<sup>®</sup> database for that parameter
- Customers want to see how their estimates compare to analogous systems
- Customers want to be able to audit their estimate in order to present it forward to management and customers
- PRICE<sup>®</sup> identified the need to be able to offer Findings that would support these customer needs
- To help users trust the TruePlanning<sup>®</sup> technology, we need to better explain the mechanics of the CERs and the nature of the underlying data used to develop the models

# Project Goals

- Create datasets that allow customers to have perspective of how their input values relate to PRICE<sup>®</sup> research
- Create datasets that allow customers to compare their estimate to analogous projects
- Create datasets to provide context to support estimates for new projects
- Provide datasets to the cost community to support data driven estimating
- Integrate these datasets into the TruePlanning<sup>®</sup> framework to provide visualizations and comparisons to support interpretation and understanding of the estimator



# Progress – Data Collection

- There are four general sources for data
  1. Publicly available data such as :
    - *Selected Acquisition Reports (SARs)*
    - *Government Accountability Office (GAO) Reports*
    - *US Bureau of Labor and Statistics*
    - *Academic papers and publications*
  2. Purchased Data such as:
    - *International Software Benchmark Standards Group (ISBSG) databases*
    - *Janes*
    - *Haystacks*
  3. Proprietary data shared by organizations (aggregated or as metadata)
  4. Government data sources such as EVM Reports and other government databases

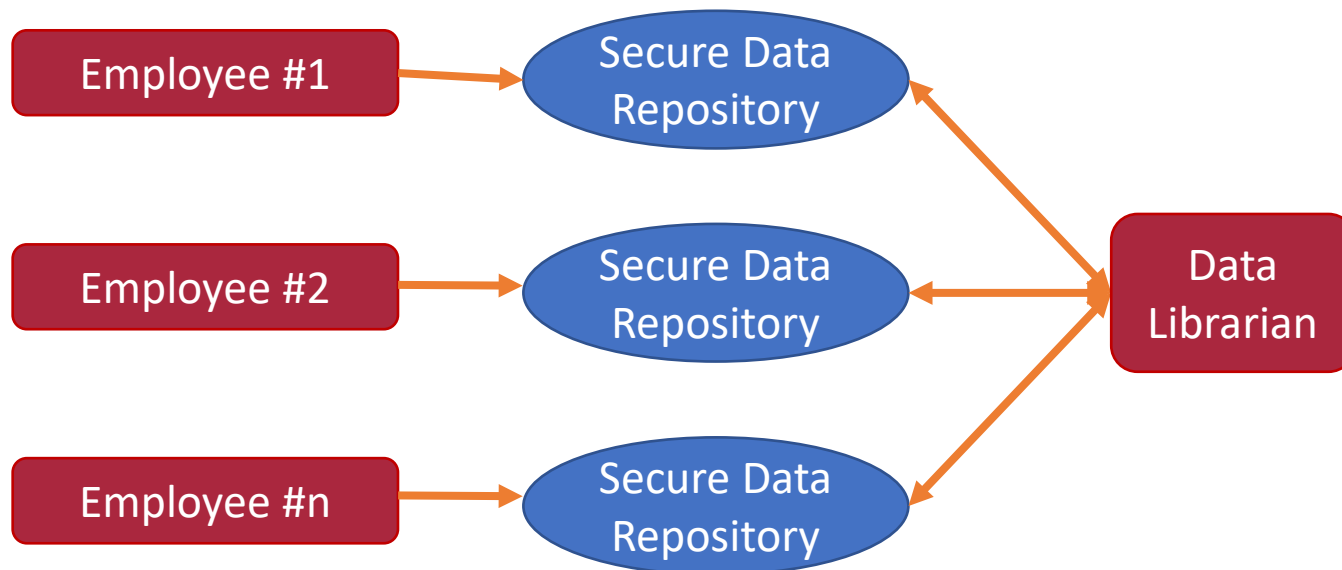


# Progress – Sources Inspected

- Software and Hardware Datasets that have been or can be *calibrated* in TruePlanning®
  - PRICE® Hardware Knowledge Network
  - Haystack Database
  - Space Component database
  - GSA Advantage
  - UK Ministry of Defense (MoD) public database
  - International Software Benchmarks Standards Group (ISBSG) Database
  - PRICE® Software Knowledge Network
  - Software data points culled from academic and conference papers (28 papers with potential ~1500 software projects)

# Progress – Data Security Plan

- **Data Availability/Sensitivity**
  - Use “Open-Source Intelligence” (OSINT) aka open-source data
  - Respect proprietary nature of data source where applicable
  - Use encrypted system or file directory (e.g. VeraCrypt)



# Data Normalization

## Secure Data Repository



### Data Transformation Tasks

Initial Mapping to current database columns

Remove blank, useless columns/datapoints

Check bounds for TruePlanning/Verify data type

Normalize numerical/text data

Repeat Check bounds/verify data type

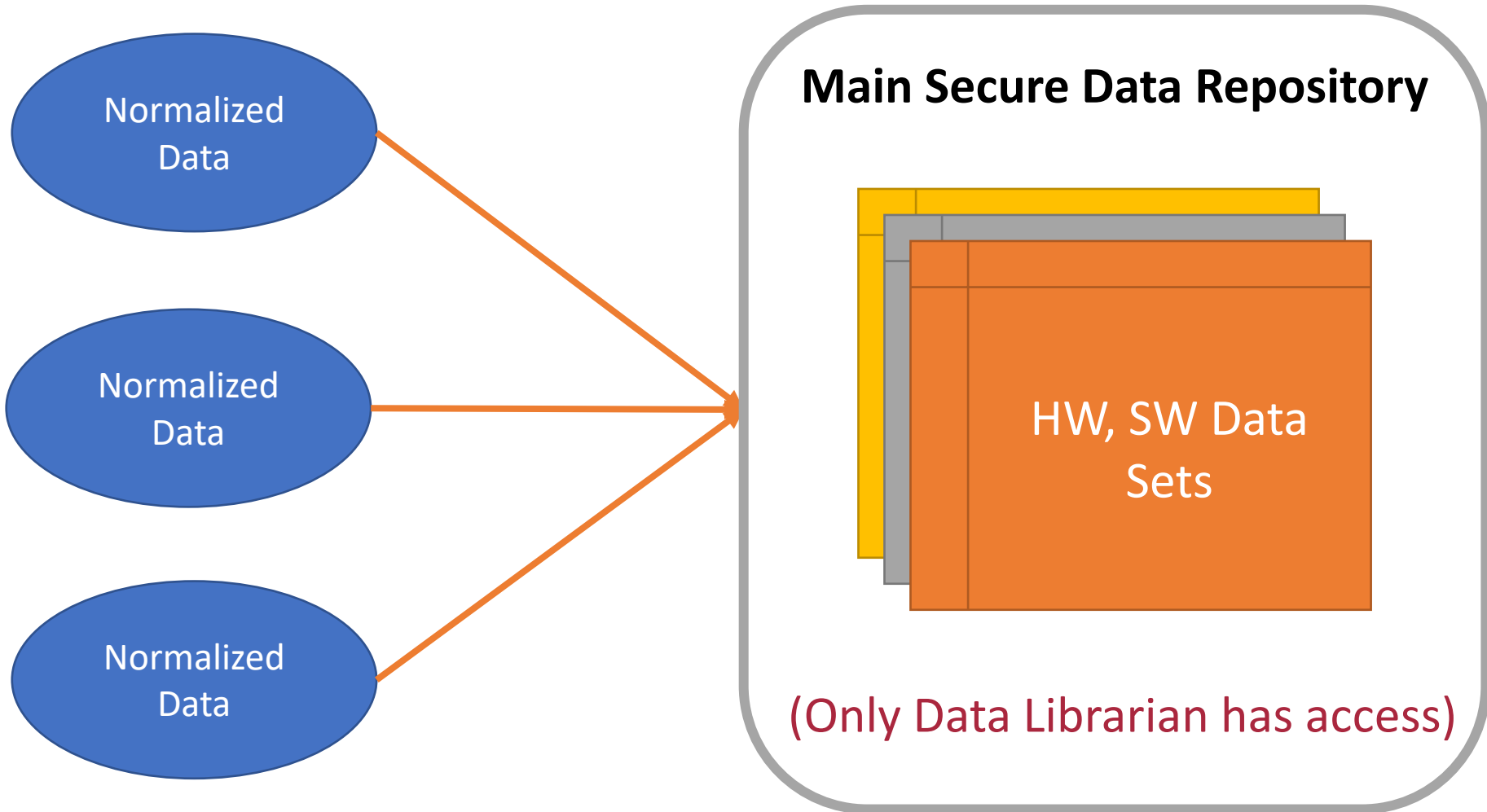
Assign ID Number, Hardware/Software Category

Calibrate (if needed)

# Calibration Details

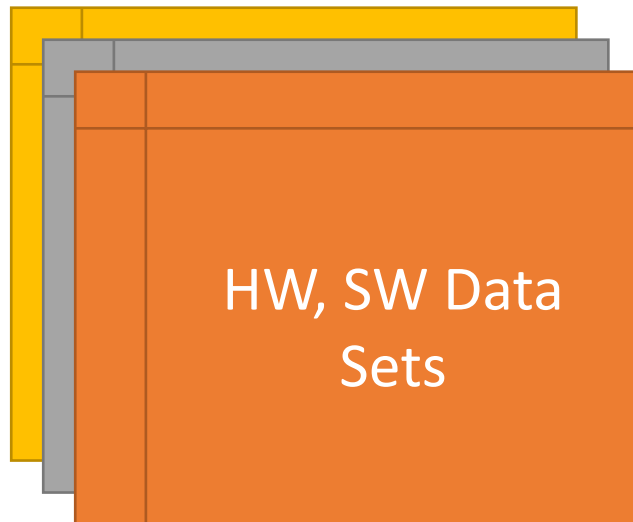
- Calibration: optimize Cost Object costs based on a chosen input value
- Target inputs are based on component type:
  - Hardware: MCPLXS/MCPLXE
  - Software: Organizational Productivity
- PRICE<sup>®</sup> Research team is working on standardizing methodology for future Data Initiative calibrations

# Data to Main Repository



# Splitting Data into 2 data Sets

## Main Secure Data Repository



**1. Proprietary and Open-Source Findings Data –**  
Aggregated data for high level use/guidance

**2. Open-Source Data Sets –**  
Component data available for general use

## Progress – Expert Judgement Elicitations

- Data is king... but often data can not be found or harvested efficiently
- It is often necessary to supplement sparse data with Subject Matter Expert (SME) judgement

*“The knowledge held by expert practitioners is too valuable to be ignored. But only when thorough methods are applied, can the application of expert knowledge be as valid as the use of empirical data. The responsibility for effective and rigorous use of expert knowledge lies with researchers.”*

Drescher, M. et. Al, “Toward Rigorous use of Expert Knowledge in Ecological Research. *Ecosphere* 4, 1–26, 2013



## Progress – Expert Judgement Elicitations

- Standard process is essential to successfully elicit expert opinions while tempering bias
- The process in place at PRICE<sup>®</sup> includes:
  - Select a facilitator
  - Select a group of experts (4-5) based on their area of expertise and experience
  - Communicate the purpose of the elicitation to the experts
  - Use anonymous surveys to solicit feedback on the topic
  - Distribute the results along with rationale and allow experts to update their advice based on others rationale
  - After several rounds of anonymous surveys – get the experts face to face to reach consensus on areas of dispute



# Live Demo

## Results to Date

Data Type	Approx. Number of Datapoints	Cost Object Findings Generated
Hardware	9,900	500
Software	3,000	130



# Summary

## Next Steps

- Collect more data
- Refine Technical data
- Optimize search capability for data and Cost Object Findings
- Work with Product Development and PRICE® TruePlanning® customers to refine the integration of PRICE® datasets with program features.

# Contact PRICE®

[www.pricesystems.com](http://www.pricesystems.com)

[www.modelbasedcostengineering.com](http://www.modelbasedcostengineering.com)

1-800-43PRICE

[Arlene.Minkiewicz@pricesystems.com](mailto:Arlene.Minkiewicz@pricesystems.com)

[Cara.Cuiule@pricesystems.com](mailto:Cara.Cuiule@pricesystems.com)



TrueExplorer



TrueFindings



PRICE® Models



TruePlanner



TrueMapper



TrueBOE



TrueXLS

Search &  
extract data  
from the PCA  
Ecosystem

Manage &  
Analyze Data  
Sets

Predictive  
Models

Integration  
Framework

Customer  
Data Mapping

Basis-of-  
Estimate  
Generator

Access PCA  
Engine  
from Excel