TCM-3756

Recommended Practice for Project Historical Database Development

John K. Hollmann, PE CCP CEP DRMP FAACE Hon. Life and Peter R. Bredehoeft, Jr. CEP FAACE

Abstract

The value of good project historical data and metrics for use in project planning (estimating, planning and scheduling, risk analysis, etc.) has always been recognized by cost engineers. The AACE® virtual library includes examples of successful implementations. However, the challenges of developing and maintaining a database (sustained demands on resources and budgets and extended time to achieve objectives) have resulted in relatively few companies successfully implementing them. However, increasing interest in artificial intelligence (AI), analytics, business intelligence tools and the availability of commercial software has raised interest. This interest resulted in the development of Recommended Practice (RP) 114R-20 "Project Historical Database Development". The RP is a guideline for requirements assessment, specification, development, implementation, and maintenance of a project historical database system. A database maturity model is also included. The intended audience is owner, contractor and agency organizations having access to project estimate and/or actual data. The main focus is on databases for estimating, planning and scheduling, and risk management uses (e.g., estimate validation, conceptual estimating and scheduling, parametric risk modeling, etc.). However, databases may also support resource planning, project system benchmarking and performance improvement, forensic analysis, and other processes in the TCM Framework.

AACE INTERNATIONAL RECOMMENDED PRACTICE

114R-20

PROJECT HISTORICAL DATABASE DEVELOPMENT

AACE INTERNATIONAL

AACE® International Recommended Practice No. 114R-20

PROJECT HISTORICAL DATABASE DEVELOPMENT

TCM Framework: 6.3 – Asset Historical Database Management
10.4 – Project Historical Database Management

Rev. April 1, 2021

Note: As AACE International Recommended Practices evolve over time, please refer to web.aacei.org for the latest revisions.

Any terms found in AACE Recommended Practice 10S-90, *Cost Engineering Terminology*, supersede terms defined in other AACE work products, including but not limited to, other recommended practices, the *Total Cost Management Framework*, and *Skills & Knowledge of Cost Engineering*.

Contributors:

Disclaimer: The content provided by the contributors to this recommended practice is their own and does not necessarily reflect that of their employers, unless otherwise stated.

Peter R. Bredehoeft, Jr. CEP FAACE (Primary Contributor)
John K. Hollmann, PE CCP CEP DRMP FAACE Hon. Life (Primary Contributor)
Fred J. Wellman (Primary Contributor)
Christopher W. Carson, CEP DRMP PSP FAACE
Larry R. Dysert, CCP CEP DRMP FAACE Hon. Life
Bruce G. Elliott, CCP FAACE
William G. Hombach
Dave Kyle, CCP CEP
Michael Lesnie
Arvid Markus
Dan Melamed, CCP EVP FAACE
Nick Papadopoulos
Todd W. Pickett, CCP CEP FAACE
Avi Sharma
H. Lance Stephenson, CCP FAACE
Subhash Tuladhar

April 1, 2021

TABLE OF CONTENTS

April 1, 2021

1. INTRODUCTION

1.1. Scope

This recommended practice (RP) of AACE® International defines the basic elements of and provides broad guidelines for evaluating, developing and maintaining project historical data management systems (i.e., a database). The intended audience is any organization considering a new implementation or a significant improvement in capital project historical database maturity (a database maturity model is included). Some organizations have integrated processes and data in place, and as such, an historical database may be a relatively small step; the RP's length is driven by the needs of organizations with less mature information frameworks or in need of significant improvement. Use the parts of the RP that apply to one's situation.

Database implementation and improvements are accomplished by executing process improvement projects. However, this RP is not about how to execute such a project. Instead, it describes the main elements that should be considered in defining the database project scope including considering the database life cycle operation and management. It provides a basis or a framework for planning.

This industry-generic RP is aligned with the *Total Cost Management (TCM) Framework* [1]. In TCM, asset and project historical database management processes are covered in Chapters 6.3 and 10.4 respectively. Every process map in TCM connects to the database processes because learning and using information from history is critical to every strategic asset management and project control process and function. In addition, this RP addresses the needs of owners, contractors, consultants, agencies and others with a myriad of potential uses for project estimated and/or actual data.

This RP focuses on supporting use in project cost estimating, schedule planning and development, and risk management. However, databases may also support resource planning, project system benchmarking and performance improvement, forensic analysis, and other processes in TCM. Figure 1 from TCM Chapter 10.4 shows the typical information flows as well as the general planning methods and tools uses of information from a project historical database.

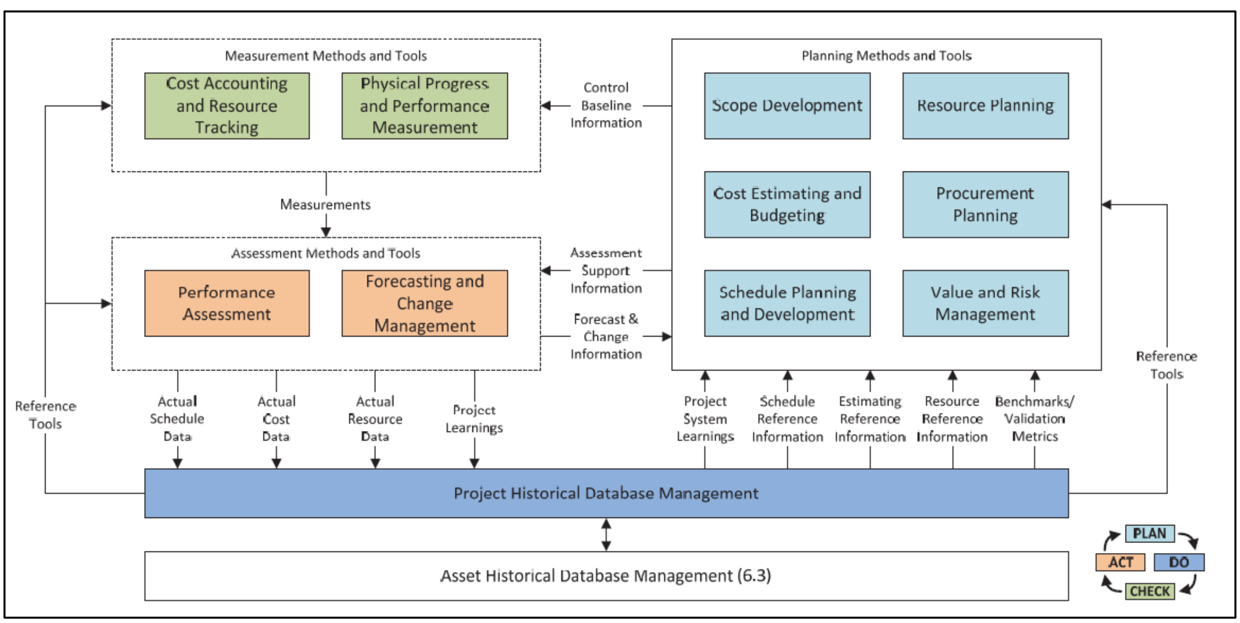

Figure 1: TCM Project Control Information Flow Supported by a Database [1]

Industry is making significant strides to integrate processes and data flow across the life cycle of projects, between systems, and among project participants. However, fully integrated life cycle systems, from the earliest estimates to final actuals and beyond, are not yet the reality for most organizations and project systems. Therefore, this RP covers everything from spreadsheet-based to near plug-in ready commercial applications. Building information management (BIM), enterprise resource planning (ERP), and cost management systems, addressed in this RP at a strategic level, perhaps represent the most integrated processes and systems in current practice. These are aided by visualization and business intelligence software. However, these do not address all the requirements of and uses for a project historical database. In this RP, the term database refers to a project historical database (unless specifically identified otherwise).

1.1.1. RP Exclusions

- The RP does not define specific activities of a database implementation or improvement project; only considerations as a basis for planning its scope; i.e., a framework for planning.
- The RP supports but does not include end use practices that have their own RPs such as estimate validation, estimating system database development, parametric modeling of cost and risk, etc.
- The RP assumes a database supports but does not include business intelligence and advanced analytics application methods such as artificial intelligence, machine learning, etc.
- The RP is not focused on specific information technology (IT), but what a database project manager needs to know in order to work and communicate with an IT department.
- The RP excludes details of BIM, ERP and cost management systems, but covers strategic considerations for alignment with them.
- The RP excludes data on operating and maintaining assets over their life cycle (re: TCM chapters 5 and 6) other than measures of operability or other data used for project system benchmarking.

1.2. Purpose

This RP is intended to provide guidelines (i.e., not a standard) for what to consider in evaluating, developing, maintaining and applying project historical data management systems. This provides a framework and considerations for planning a database asset and project, but not a guideline for how to manage any specific database project. General software types are described but not specific software products. Most practitioners would consider these guidelines as good and reliable practices for consideration where applicable.

1.3. Background

1.3.1. Terminology

The following are key terms to understand prior to further discussion. Most of the definitions are from Recommended Practice 10S-90, *Cost Engineering Terminology* [2]; the terms and definitions in *italics* are not.

- *ALLOCATION – In historical database management, the process of distributing or assigning estimate or actual cost as captured to specific accounts in the database structure.*
- *ANALYTICS – Inferential statistical methods (e.g., regression) applied to understand metric relationships and behavior (e.g., trends over time, trends by project attribute, relationships between metrics, etc.).*
- *BACKFILL DATA – Legacy data used to augment a dataset when sufficient current information is not available.*
- *BACKUP – Supporting documents for an estimate or schedule including detailed calculations, descriptions of*

April 1, 2021

data sources, and comments on the quality of data.

- BENCHMARKING – A measurement and analysis process that compares practices, processes, and relevant measures to those of a selected basis of comparison (i.e., the benchmark) with the goal of improving performance. The comparison basis includes internal or external competitive or best practices, processes or measures. Examples of measures include estimated costs, actual costs, schedule durations, resource quantities, etc.
- *BUSINESS INTELLIGENCE – Tools to transform raw data into useful information, typically supported by special purpose software that includes DATA VISUALIZATION.*
- *DATA AND INFORMATION – Data are facts such as attributes, cost, time, quantities, and so on. Information is data in context or in relation to other data such as metrics.*
- *DATA CLEANING – Correcting data errors and other discrepancies. Ensuring that attributes, labeling, and categorization of data is correct.*
- *DATA VISUALIZATION – Graphical representation of information and data, typically supported by special purpose software. Tools with analytical capability are referred to as BUSINESS INTELLIGENCE software.*
- *DATABASE (GENERAL) – The repository of data and information. May be a relational database with multiple tables of records with key identifiers and fields (a spreadsheet is a table). Operation uses structured query language (SQL). Term may be referring to a "database system" which includes functionality such as normalization, analytics, and reporting.*
- DATABASE (HISTORICAL) – Records accumulating past project experience stored as data for use in planning, estimating, forecasting and predicting future events. Often includes data that has been processed so as to facilitate planning and other purposes such as validation and benchmarking (e.g., metrics, etc.).
- *DATASET/SAMPLE – A grouping or subset of data that is selected based on specific criteria. Sometimes referred to as a comparison dataset (or compset in some vernacular).*
- ESCALATION - A provision in costs or prices for uncertain changes in technical, economic, and market conditions over time. Inflation (or deflation) is a component of escalation.
- ESTIMATE VALIDATION – (1) A quality assurance process, typically quantitative in nature, to test or assure that an estimate of cost or time meets the project objectives and strategy in regard to its appropriateness and purpose (which may include competitiveness or other organizational strategies identified for the estimate). (2) A form of benchmarking that compares relevant estimate cost, time and/or resource measures (e.g., metric ratios) to those of a selected basis of comparison.
- *GO-BY – In historical database management, something captured (e.g., report, estimate, schedule, model, etc.) that can be used as an example or template to consider or follow.*
- *METRICS – In historical database management, ratios of cost, duration, quantities and other resources that are used for estimate validation, conceptual estimating, performance evaluation and other purposes. Usually not stored in the database as such but calculated on demand as needed after normalizing the ratio elements.*
- NORMALIZATION − In database management, a process used to modify data so that it conforms to a standard or norm (e.g., conform to a common basis in time, currency, location, etc.).
- PRICE INDEX – A number which relates the price of an item at a specific time to the corresponding price at some specified time in the past.
- *RULE OF THUMB – A measure or value that is based upon experience or historical data for typical elements of work. These values may or may not have been based on significant analysis and have generally not been adjusted for specific project scope and/or normalization. If the value is important to a decision or end use, it's quality should be tested or checked.*
- SCOPE DEFINITION – Division of the major deliverables into smaller, more manageable components to: 1) Improve the accuracy of cost, time, and resource estimates; 2) Define a baseline for performance measurement and control; and 3) Facilitate clear responsibility assignments. *In respect to a project historical database, scope definition relates to the organization and division of content of data records to facilitate queries and data and metric comparisons.*

1.3.2. Early History of Project Historical Databases

A fundamental practice of cost engineering is the use of data from past projects for use in planning future ones. For example, a 1960 AACE paper on the "Essentials of Cost Control" spells out the need to produce a final cost analysis upon project closeout. It states that the "information is necessary for estimates and cost projections of future projects" [3]. By the 1970s this data was increasing being computerized. A 1984 Cost Engineering article summarized developments and stated "Cost data, organized on a computer system and poised vis-à-vis the next estimate, is the goal if not the reality of the estimator" [4]. In the 1980s interest grew to more integrated databases to capture more than just cost, but integrated data from all functions of the project [5]. In 1987, a National Research Council study laid out a conceptual framework and rules for "Integrated Data Base Development for the Building Industry" [6]. In 1995, an AACE paper described an owner company's custom programmed project historical database system as addressed by this RP; i.e., it was designed to support all the cost engineering planning functions in TCM [7]. Now, commercial, cloud-based systems are available, and visualization and business intelligence applications help tap into sometimes disparate databases. Efforts are increasingly focused on integration of processes and data across the capital project management spectrum, but many still have a long journey to that destination.

1.3.3. Principles of Database Development and Management

The history above provides learnings and principles to guide practices. The first principle is that developing a database can be a significant project (i.e., the TCM Framework applies). The TCM process, and database development, start with traditional good project practices such as establishing and communicating clear objectives and applying strong leadership and team development. TCM calls for scope development and investment decisions to be done using a phase-gate project system.

However, beyond general TCM and project management principles, experience provides other rules and lessons. Some are from the historical references above (e.g., the National Research Council report), but also from the RP contributors who have been part of database development projects and managed databases of various types and maturities. The following rules and lessons, dispersed in the recommended practice sections, are summarized here to reinforce important points:

- Expectations: Good systems take time to develop, populate and achieve objectives, so set realistic expectations with management; don't overshoot or overpromise (but do strive for early wins; see backfill). On the other hand, some of the references to this RP will help explain the potential benefits.
- Culture: a quote often attributed to Peter Drucker is that "culture eats strategy for breakfast". Implementing a database touches many elements of an organization and their processes; make sure your database implementation strategy considers the culture.
- Stakeholders: a TCM-capable system may have multiple stakeholders; business planning, estimating, planning and scheduling, project control, risk management, procurement, finance and accounting, design, and even legal (forensics). Get their input and involve them as appropriate. However, avoid encumbering the system with capabilities outside the objectives.
- Business Process Synergies: Project historical databases stand at the crossroads of various business processes requiring focus on process interactions to facilitate data flow and usage.
- Backfill/Early Wins: A database without data is of no value. The project is not complete until it is producing useful product (early wins). Plan on a significant effort to process and input past project records (until then, make sure project files are not discarded).
- Process Before Software: Understand and define the database management process (and the rules here), and get the groundwork rolling (e.g., structure, forms, etc.) before diving into software implementation (however, do consider software capabilities in defining a process).

- Structure: The success of the development in large part depends on the supporting coding structures (i.e., asset breakdown and code of accounts). The better that structure is defined and standardized throughout the organization, and aligned with industry practice, the better the value obtained from the system.
- Scope Attributes: Using data and metrics rely on like-for-like comparisons[1]; structure the data and capture asset and project attributes in a way that facilitates queries and results in valid comparison samples (e.g., capture not only entire project data, but break it down by meaningful scope elements).
- Data Quality: Define minimum data quality needs upfront that will support the kind of products and statistical confidence (e.g., metrics) to meet the system objectives.
- Data Sourcing: Plan for dealing with disparate sourcing such as contractors, distant teams, temporary hires, accounting, etc.; establish and enforce contract requirements from 3rd parties, and train employees and contract hires as needed.
- Derivative Information (e.g., metrics): If a measure can be created from captured data, don't capture it; calculate it in the system (e.g., don't capture unit costs; capture quantity and costs). Similarly, don't capture end use data such as *estimating system data*; use the historical data to derive and or calibrate it.
- Capture data at the source: The person(s) capturing the data (and often allocating it) should be someone with the most knowledge about the project and the particular data.
- Expiration: Without good quality normalization, data reliability and usefulness will decline rapidly over time (understand and apply escalation, currency and location adjustment best practices)
- Resourcing: Database development projects are usually not a spare-time endeavor. It may be part-time, but be realistic in resource planning, and make sure the right skills and knowledge are at hand or acquired.
- Skills and Knowledge: A degree of estimating, control, planning and scheduling, risk and statistical skills and knowledge are required to make the process work; plan on it.
- Sustainability/Scalability: Do not create something that is difficult to manage. Consider evolving maturity and scalability (i.e., a system that can grow with the data and evolving uses and users).

In addition, the following are considerations to better frame planning efforts:

- Estimating databases (including durations) differ from project databases. Estimating databases support estimating systems (e.g., unit hours, production rates, etc.). Project historical databases typically support planning (include risk information), estimate and schedule validation, and to some extent conceptual estimating and scheduling. Typically, project data is analyzed to generate metrics which are typically ratios of various types (unit hours and material costs can be metrics).
- ERP systems are not project historical databases. They have a database(s); however, much if not most of the data will not be structured in a way that is useful for project planning (e.g., captured by contract, but not by discipline). That is not to say these systems, including procurement data, are not a good source of some data.
- Allocation required. A premise for project databases is that "there is no such thing as actual data". This means that raw data is often not clean or organized in a way that is useful for database purposes. Optimally, only quality checks of raw data will be required, but be prepared to resource the task of cleaning, allocating or otherwise processing actual data into a usable form (see ERP comment), using reliable references to guide the allocation (e.g., original estimate breakouts, change records, etc.)
- Allocation/processing data for planning use also means historical database data is not accounting data; if one advertises data as auditable, be prepared for legal and other implications (e.g., discovery, retention constraints, etc.).
- Another premise is "the more you ask for, the less you get". A graph of extent of requested data input (x-axis) to quality of data obtained (y-axis) is an inverted U. As more detailed is requested, the difficulty of providing it increases to a point where the provider may cut corners (leaves blanks, makes gross

[1] Sometimes also referred to as an "apples-to-apples comparison".

April 1, 2021

assumptions, etc.) or the system otherwise breaks down. Another example is that contractors may balk at data that exposes margins. Find the optimum balance.

- The concept of critical mass applies. A small amount of data is not much better than no data. Until a system achieves a robust volume and variety of applicable, appropriately detailed, well maintained data and metrics, it will not satisfactorily achieve its purpose. This usually requires initial database backfilling with cleaned/normalized past records (rule-of-thumb; it takes a minimum of 5-10 applicable, quality observations before a metric is statistically significant).
- The best output information is the product of analysis. While a system can calculate a metric, that is often just the start. Getting the maximum value from data requires some modest level of statistical competency in the organization. This includes understanding descriptive (mean, mode, median, confidence) and inferential (regression) statistics. This is the entry level to analytics (e.g., AI, machine learning).
- Good graphics needed. Graphical representation of metrics, trends, and so on greatly enhance understanding. This is an important feature/benefit of commercial software; however, visualization and/or business intelligence software is a great help with any data source.
- Security is critical. Project historical data is usually commercially sensitive for owners and contractors; therefore, security is always an issue to address. If collaboration with external parties is a goal of the database project, this will require particular attention as to what can be shared.
- Data takes various forms. To support project planning, consider capturing (or linking to) various documents and files such as maps, photos, key drawings, reports, analyses, bids and quotes, and so on.

1.3.4. Introduction to Data Applications and Uses

An early step in defining database scope is defining its end uses. One of the early uses of project historical data was to support estimate preparation and/or estimate database development. The range of uses has increased over time to support scheduling, risk analysis and other uses. This RP envisions the database supporting many of the TCM Framework processes as shown previously in Figure 1.

Table 1 is an outline of the range of potential database system uses categorized by TCM sub-process. It is not all-inclusive. It is recognized that historically cost estimating uses have often driven initial database planning; however, one goal of this RP is to make sure other potential uses are considered as well.

April 1, 2021

TCM Process/Function	Potential Database System Use
Asset Planning	• All planning uses, but at a conceptual level • Information for scope planning (quantities, attributes, photos, maps, etc.)
Investment Decision Making	• Information to support comparison of alternatives • Data to support study of past decision quality
Project Implementation	• Assess and benchmark performance of project system and organization
Project Scope and Execution Strategy Development	• Quantity reference data and metrics for planning and validation • Work breakdown structure templates • Execution strategy templates
Schedule Planning and Development	• Duration reference data and metrics for planning and schedule validation • Schedule templates
Cost Estimating and Budgeting	• Estimate reference data and metrics for estimate validation • Resource to support estimating system database development or calibration. Data and metrics for estimate preparation and modeling • Past estimates and actuals for analogy estimating • Support special studies (past escalation, location factors, productivity, rates)
Resource Planning	• Study resource usage, availability, limitations • Resourcing strategy templates
Value Analysis and Engineering	• Reference data and metrics for value analysis/value engineering and other value improving practices analyses
Risk Management	• Data and metrics to develop and calibrate empirically-based risk quantification methods and tools • Risk identification, risk treatment and risk response templates and checklists
Procurement Planning	• Study effectiveness of procurement strategies and practices • Study price trends and behavior • Study supplier bidding behavior • Data to support strategic purchasing • Contracting and procurement and specification templates
Project Control Planning	• Lessons learned for control planning • Control plan templates
Project Performance Assessment	• Periodic or key milestone metric comparisons and time trends • Benchmarking of project performance during execution • Lessons learned regarding performance
Forecasting	• Data and metrics to support change impact and risk estimates for cost and schedule • Analysis of performance trends and effectiveness
Change Management	• Data and metrics to support change impact and risk estimates for cost/schedule • Change log templates
Forensic Performance Assessment	• Data and metrics to support claims analyses • Lessons learned for claims avoidance/analyses

Table 1: Potential Database Uses in Support of Various TCM Processes

April 1, 2021

1.3.5. Specific Use Cases

Research for this RP surfaced several typical use cases; i.e., case studies of database developments to meet specific needs. These are summarized in Table 2. They are not prescriptive but suggest typical practices (e.g., it presumes few owners do detailed, control level estimating or scheduling in-house).

User	Function	Typical Data Uses
Owner	Estimating and Scheduling	Conceptual estimating/scheduling metrics, analogs and reference information, metrics to validate 3rd party estimates/schedules and bids, data for parametric modeling, compare to external benchmarks, performance measurements (including during the course of the project; not just beginning or end), location and site studies
	Risk	Parametric risk modeling, risk identification guides, cost growth/schedule slip outcome analysis (distributions, bias, etc.)
Contractor	Estimating, Scheduling, Bidding	Detailed estimating database development/calibration, production rates, productivity analysis, unit price development, conceptual estimating/scheduling metrics, analogs and reference information, data for parametric modeling, performance measurements including during the course of the project, margin analysis, and claims support. Also, sales/marketing (business development) support
	Risk	Risk identification guides, margin outcome analysis (distributions, etc.)
Consultant	Estimating and Scheduling	Metrics to validate client estimates/schedules, client performance measurements. Support project feasibility stage planning.
	Risk	Risk identification guides, cost growth/schedule slip/margin outcome analysis (distributions, etc.)

Table 2: Typical Database Use Cases

1.4. Database Process Maturity

To help frame phased database planning efforts, and later performance assessments, Table 3 outlines database process maturity levels that have been observed. This is not a measure of the overall project system maturity, but the project system and databases often evolve together. Note that different parts of an organization or project system may be at different levels of maturity.

April 1, 2021

Maturity	Description of Project Historical Database Status
Ad-hoc	• Process includes maintaining study, estimate and actual file records which are assessed on occasion or as needed in special efforts from various repositories to extract high level metrics. • Code of accounts are usually not standardized, there is no formal project cost and schedule closeout procedure or forms (or they are not used effectively), and contract terms do not require contractor adherence to data norms. • There may be some attempt to capture data in a standard form, but efforts are constrained by a lack of project system and/or control discipline and/or consistency between businesses or regions, and/or in staffing and resources. • Estimate/schedule validation is not done with any rigor. • Often executed by individuals who create their own (personal) historical reference data.
Managed	• Processes have crossed the threshold from capturing data in files and forms, to maintaining that data in a repository ranging from a spreadsheet(s) to a relational database. • There is at least a part-time resource with responsibility for managing the data and process such as it is (often with estimating oversight responsibilities as well). • Code of accounts and project system and/or control process discipline and standardization is often still marginal and team support of the effort inconsistent. • The process may be limited to isolated business or regional units. • The company may be involved in industry benchmarking. • Estimate/schedule validation is done for strategic projects as special efforts.
Robust	• Processes have developed custom programmed or implemented commercial databases. • Code of accounts may not be standard, but ability of teams to translate accounts is established. • The project system and control process incorporate steps to capture data, and contract terms are established to obtain structured data from contractors. • Usually, the organization has inhouse estimating competency at some level. • Likely involved in industry benchmarking. • Rigorous estimate/schedule validation is done for all major projects.
Optimized	• Processes are robust, plus they have evolved to support many advanced uses such as conceptual estimating, empirical risk modeling, resource planning, project process improvement and other functions within the TCM Framework. • Code of accounts are standard throughout the organization, and projects of most sizes and types are captured and analyzed, as well as data from intermediate project milestones. • Database management is staffed on a full-time basis and is recognized as a company core competency. • The database is considered a vital element of the company's capital project system and competitiveness in investment or bidding decisions. • Company system integration (BIM/ERP/EVM, etc.) facilitates elements of database operation. • Likely a leader in industry benchmarking efforts. • Validation uses advanced analytics.

Table 3: Project Historical Database Process Maturity Levels

Database maturity is generally driven by the need to benchmark capital investments, estimates, and schedules. There are typical capacity thresholds seen at which the capital or bidding demand is such that database efforts cross from Ad-hoc to Managed and then to Robust. The threshold is defined roughly by the number of estimates and project completions per year, how diverse and volatile these numbers are, and the availability of in-house resources and overhead budget to develop and sustain a managed database. Progressing to the Optimized level is more a matter of the capability of the application rather than the amount of available data. Companies usually transition through the maturity stages; however, some will take the leap from Ad-hoc to Robust. Note that cyclical capex spending/bidding and staffing hinders maturity development and often results in maturity regressing from Managed back to Ad-hoc status; Robust systems are more apt to be sustained.

2. RECOMMENDED PRACTICE

This RP describes special considerations for defining project historical database *asset and implementation project scope* including considerations for database life cycle operation and management. This description is organized by topics typically considered in a scoping or feasibility study (i.e., definition to support development of a Class 5 or 4 project estimate and schedule.) With this basis information, a project team should be able to develop an estimate and schedule for implementing or improving a database using basic TCM and AACE recommended practices.

2.1. Database Management Process

TCM chapters 6.3 and 10.4 document typical historical database management processes, including process maps; those chapters should be reviewed. TCM chapter 11.3 on information management should also be reviewed. Further, each TCM sub-process (e.g., *7.3 Cost Estimating and Budgeting*) has a map showing inputs and outputs to and from the database process. Finally, the *Introduction to Data Applications and Uses* section in this RP lists typical (but not all-inclusive) database uses to consider in the database design. Figure 2 shows how one company defined their database process for management review [8].

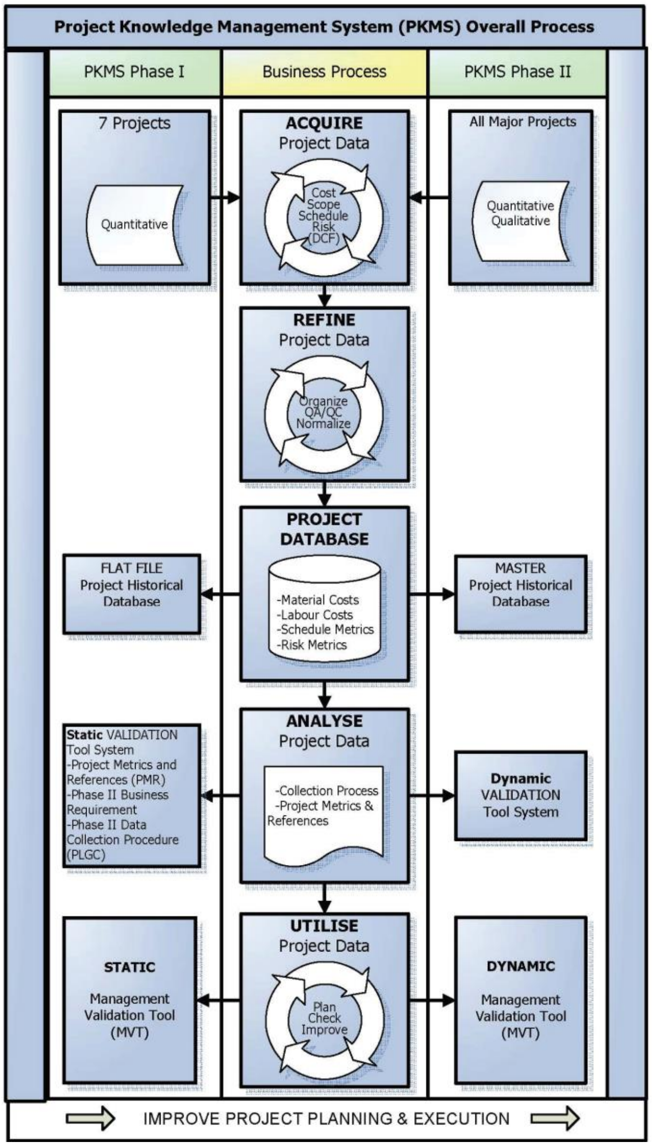

Figure 2. Example Database Process and Phasing [8]

2.2. Stakeholder Management

A project historical database may have multiple stakeholders. These may include: business planning, estimating, planning and scheduling, project control, risk management, procurement and contracting, finance and accounting, engineering and design, IT and even legal (forensics) and HR (resourcing studies), get their input and involve them as appropriate. However, avoid encumbering the system with capabilities outside the objectives. For example, finance may see the database as a means of determining values for asset capitalization; however, in doing so, this may hinder capturing cost in a way that is optimal for project planning (and also make the system auditable which creates complications). The *Introduction to Data Applications and Uses* section of this RP outlines why a party may be a stakeholder in respect to applications; however, their role may only be as a *source* of data and information. A key stakeholder will always be the project control function; not only as a potential user (e.g., metrics for change evaluation or mid-stream performance assessment), but as the primary provider of inputs via the project closeout process. Similarly, contracting is central to assuring that suppliers provide data as needed.

However, the ultimate stakeholder is business management making investment or bidding decisions, and for whom better information results in better decisions. A common dilemma is that demand for a database comes from the bottom of the organization (i.e., estimating). But the funds for the database project must be approved at the top. The project leader must ensure that the decision makers understand the value of the database process and products throughout the TCM spectrum of project system functionality. The more support from the executives, the more likely the implementation will be a success. Figure 3 provides an example database information flow diagram that can be used to help identify stakeholders to the database system [7].

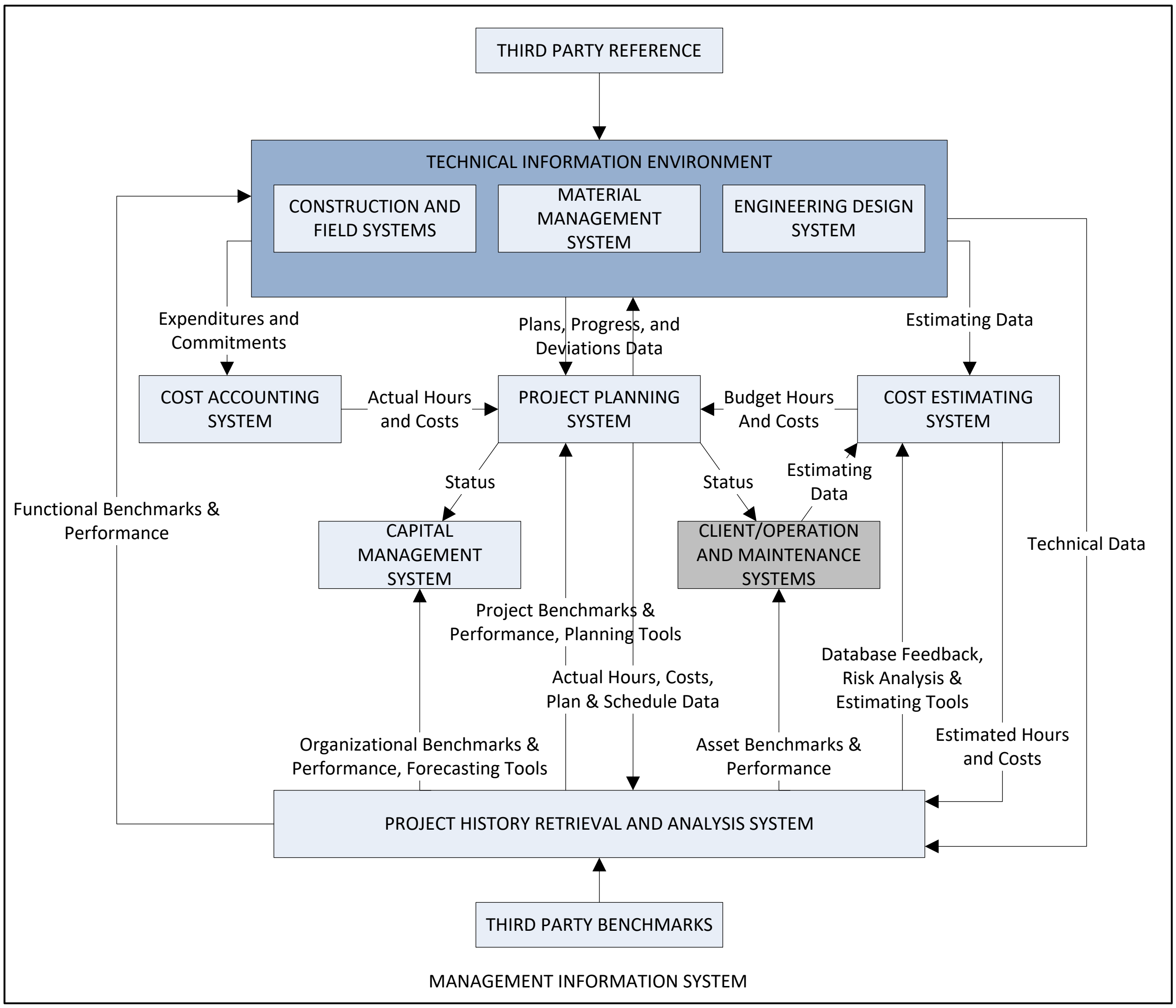

Figure 3. Example Database System Diagram to Support Stakeholder Management [7]

2.3. Needs/Requirements

2.3.1. General Requirements

Requirements elicitation and analysis is the entry process to TCM (chapter 3.1) and given the wide range of stakeholders in a database, it is important to do this well. Per TCM, "*It is a process of identifying stakeholders and their needs, wants, and expectations; probing deeper into them; and documenting them in a form that supports planning, communication, implementation, measurement, and assessment*". As mentioned, it is typical for the estimating function to make the first call for a database to support estimate and bid preparation and/or validation. However, it is often other functions that stand in the way of successful implementation because their needs were not considered or may differ.

For example, estimating may devise an extremely detailed data collection form and database structure, but if this is beyond the capability of project controls to support, and contractors refuse to provide such detail, it will likely fail. It must be fit-for-use. The more that other parties do not see a burden, but a benefit, the more likely the development will succeed. The list in the *Introduction to Data Applications and Uses* section is a good starting point for assessing potential wants and needs. A good practice is to develop some sort of a business sponsored working group of key stakeholders to advise and guide the database project team and to review development.

To support database planning, another good practice is to evaluate the practices of peer companies to see what they have done and how they did it. This RP includes a list of case-study papers for that purpose. Similarly, database software vendors may be consulted. Figure 4 is from one of those papers [7]; it shows the end uses and metric products that were identified in requirements assessment.

End Product/Use	Common Relationships	Example Calculations	Units
Rough Order of Magnitude (ROM) Cost Estimating Relationships (CERs) *Mgmt. Perf. / Quality Review* *Client Perf. / Quality Review* *Estimating Tools* *Est. Database Calibration* *Capital Mgmt. Forecasting*	Cost / Cost	DFL$ / DFM$ Total$ / Equipment$	%
	Labor-hours / Cost	DFL-hours / Equipment$ HO-hours / Total$	hours/$
	Cost / Labor-hours	DFL$ / DFL-hours HO$ / HO-hours	$/hours
	Cost / Deliverable or Output	Total Concrete$ / Total CY Total$ / Output Capacity	$/unit
	Labor-hours / Labor-hours	Process Eng.-hours / Total Eng.-hours HO-hours / DFL-hours	%
	Labor-hours / Deliverable	DFL-hours / Piece of Equipment Eng.-hours / Drawing	hours/unit
Rough Order of Magnitude Schedule Development Relationships (Cycle Time Analysis) *Mgmt. Perf. / Quality Review* *Client Perf. / Quality Review* *Planning Tools*	Time / Cost	Construction days / TFC$ Eng.-Design days / HOC$	day/$
	Time / Deliverable or Output	Debug days / No. Equip. Pieces Eng.-Design days / No. Drawings	day/unit
	Time / Labor-hours	Construction days / DFL-hours Eng.-Design days / HO-hours	day/hours
	Time / Time	Front End days / Total days Eng.-Design days / Construction days	%
Detailed Estimating Database Feedback (Productivity and Unit Rates) *Est. Database Calibration*	Actual-to-Budget-to-Estimate Labor-hours / Labor-hours	Actual-hours / Unadjusted Est. hours	%
	Same as ROM Estimating, but mostly Cost & Labor-hours / Deliverable		$/unit hours/unit
Performance and Quality Measurement (Indices and Benchmarks) *Mgmt. Perf. / Quality Review* *Client Perf. / Quality Review* *Estimating Tools* *Capital Mgmt. Forecasting* *Project Planning Tools*	Labor Efficiency: Labor-hours / Labor-hours	Actual-hours / Budget-hours	%
	Rework: Cost / Cost or Time / Time	Rework$ / Total$ Rework days / Total days	%
	Change Management: Cost / Cost or Time / Time	Non-Scope Change$ / Total$ Scope Change days / Total days	%
	Capacity Achieved: Output / Output	Actual Units / Nameplate Units	%
	Indices: Any Ratio / Process Measure	Rework% / New Process Steps Change% / FEL Index	%
	Same as ROM Est. and ROM Sched., but mostly Cost, Labor-hours, Time / Deliverable	Eng.-hours / Drawing	x/unit
Workload Forecasting Factors (CERs) *Capital Mgmt. Forecasting*	Labor-hours / Cost	Eng.-hours / Sum of Project$ Construction-hours / Sum of Project$	hours/$
	Labor-hours / Labor-hours	Electrical DFL-hours / Construction-hours Design-Drafting-hours / HO-hours	%
Risk Assessment Factors (Indices and Benchmarks) *Mgmt. Perf. / Quality Review* *Client Perf. / Quality Review* *Project Planning Tools* *Estimating Tools*	Technical Process Measures: Any Ratio / Strategic Measure	Rework% / New Process Steps	%
	Execution Strategy Measures: Any Ratio / Strategy Measure	Change% / DFL% in shutdown	%
	Location Measures: Any Ratio / Location Measure	Productivity Factor / Location Change% / Location	%
	Project Process Measures: Any Ratio / Process Measure	Change% / FEL Index	%
	Organizational/Client Measures: Any Ratio / Client Measure	Change% / selected client	%

Notes: (1) Table excludes complex algorithms developed by off-line statistical analysis, modeling, etc.
(2) DFL = direct field labor, DFM = direct field material, HO = home office, HOC = home office cost, TFC = total field cost, FEL = front-end loading

Figure 4. Database End Use and Metrics Requirements Table (Adapted from [7])

2.3.2. Requirements for a Contractor/Supplier Database

Owners are concerned with data for the entire project scope, including owner's cost. However, contractors typically focus on or are limited to narrower elements of the overall scope. Engineering focused contractors may be primarily interested in capturing their own past estimates to support internal estimate validation and to support development of feasibility estimates for their clients. Fabricators and constructors will be focused on data to support their cost models, bid or tender estimates, and associated margin determination; however, they also have access to completed project data, albeit usually not for the total scope of a project as viewed from the owner perspective. Engineering, procurement and construction (EPC) and similar contractors may have the full spectrum of estimate, bid and actual data, but excluding owner's cost.

A challenge for owners is to obtain the data for completed projects from the contractors and suppliers (i.e., actuals). It is not enough for the owner to capture bottom-line contract totals. For estimating and planning, owners need actual quantities, hours, cost and so on, as well as indirect costs, at a reasonable level of detail that contractors must provide. This may not be a challenge for reimbursable contracts but will be from lump-sum contractors and suppliers. Contractors do not want to expose their margins and will likely view cost and other resource details as competitively sensitive. These challenges can be mitigated with effective bidding/contracting practices (e.g., early contractor involvement, lump sums include cost breakouts, etc.), but that puts a premium on owners working closely with their contracting department.

A final challenge for owners is to understand the distinction of cost versus price (i.e., a unit price or lump sum from a contractor is the price to the owner, not the cost to the contractor). This is something to consider in later analysis; for example, when benchmarking a base cost estimate against historical metrics that include markups and margins, one might expect the estimate metrics to be less. On the other hand, bid data may be unbalanced (e.g., margins are loaded on items that are billed early in the project). It is possible to configure a database to capture these anomalies.

2.3.3. Structure Requirements

No database can function without a defined asset and cost structure. Optimally, the structure is the same as used in estimating, planning and scheduling, and project controls, and is common among the enterprise's organizational elements (e.g., businesses and regions). To the extent the structure aligns with other companies, contractors and benchmarking organizations, the easier it is going to be to make comparisons without translation. The structure has two main elements; an asset breakdown structure (similar to a work breakdown structure (WBS)) and a cost code of account. The topic of structure is well defined in TCM and is not repeated here; however, the following points should be considered in respect to database design.

2.3.3.1. Asset Breakdown

The term *asset* refers to the deliverable or product of the project (e.g., a building or a process plant or some part of these). Most uses of metrics require that the metric represent the scope of the project being estimated or assessed. For example, if one's estimate is for piping in a water treatment plant, one would not use piping metrics for a complex chemical process plant because these plants (i.e., assets) have widely different types and complexity of piping. For each asset type, key attribute data to support like-for-like comparison will be collected as discussed later.

Even a simple project WBS is likely to include more than one asset scope element. Therefore, a requirement for the database process is that costs and resource data from estimates and actuals be split or allocated to pre-defined asset categories such as chemical process *units* or building areas at a meaningful level of disaggregation (i.e., multiple data forms to complete). For actual costs, this disaggregation is often a challenging task requiring estimating competency. It is similar, but not the same as asset capitalization allocation (however, generally avoid using the database for capitalization). The asset categories need to be detailed enough to differentiate assets with fundamental, well-known differences, and represent the types of assets that apply to the business. Each asset category will have a key code that can be used for database queries, mainly used for selecting appropriate comparison datasets. Figure 5 is

April 1, 2021

an example of an asset/deliverable code in the International Construction Measurement Standards (ICMS) structure [9].

The example ICMS structure is at the highest level. For example, data for pipelines (code 08 in ICMS) should be captured separately for pumping stations versus the pipeline itself. Similarly, for refineries (code 12 in ICMS), metrics are usually desired at a process unit level (e.g., metrics for crude units should be separate from those for hydrotreating units).

01.	Buildings	08.	Pipelines
02.	Roads, runways and motorways	09.	Wells and boreholes
03.	Railways	10.	Power-generating plants
04.	Bridges	11.	Chemical plants
05.	Tunnels	12.	Refineries
06.	Waste water treatment works	13.	Dams and reservoirs
07.	Water treatment works	14.	Mines and quarries

Figure 5. Example Asset/Product Code from ICMS [9]

2.3.3.2. Code of Accounts

To be useful, the data must align with how cost estimates are broken out. Estimates are usually structured by a code of accounts that identifies the type of work in a consistent, hierarchical fashion. A primary account for estimating is the discipline (e.g., civil, mechanical, electrical and so on). The main challenge for databases is to decide on the level of detail required for useful metrics. For example, capturing the total hours to install piping should not be a daunting task for either estimates or actual costs, but if hours by piping alloy type (e.g., carbon versus stainless steel) were required, this could be very difficult for actual costs because construction hours are likely not tracked by alloy. One could allocate the actuals, but with some sacrifice of data and metrics reliability.

A particular challenge is how to account for direct versus indirect costs [10]. For example, an owner may be satisfied with capturing all-in field labor costs without any indirect, overheads or markup details, while a contractor, with access to detail records, may capture a breakdown of each labor cost indirect element. If project requirements for reporting indirects are unclear or vary greatly, any allocation of directs versus indirects will carry significant uncertainty (i.e., the better the requirements, the more confidence in the data).

2.3.3.3. Standard Structures

The greater the standardization of asset and code of account structures, both internal and external to a company, the better for all uses of data and metrics. There are a number of standards that have been used over the years including but not limited to:

- Buildings/Commercial/Infrastructure: CSI Masterformat® [11] or Uniformat® [12]; COA only
- Process: SCCS [13]; Asset and COA
- Process: AACE® RP 21R-98 [14]; COA only
- Mining: AACE® 103R-19 [15]; Asset and COA
- Environmental: Environmental Cost Element System (ECES) *[16]*; COA only

Several coding structures are available that cover multiple industries and asset life cycles:

- CSI OmniClass® [17]
- International Construction Measurement Standards (ICMS) [9]

Of most interest perhaps is the ICMS which is a standard being promulgated by a coalition of international associations including AACE International. Whether the ICMS is used as the primary structure of the database, it should be reviewed and considered.

While standard structures facilitate database usage, their use is often not made mandatory in a company for all projects, and they can be difficult to enforce with a diverse asset and project portfolio. Also, a standard structure may not be a good fit for a given project and force-fitting may be counterproductive to measurement performance and data quality. Therefore, to address structure diversity, a database system should have mapping functionality in place to support structure variations.

2.3.4. Asset Requirements: Attributes

Once an asset breakdown has been decided, key attributes (i.e., design data) of the asset are required to be captured. This allows for queries to select appropriate comparison datasets and to support analyses. Attributes may be either common across all project types or unique to each project type. Examples are the capacity of a plant (e.g., tonnes per day), size of a building (e.g., square meters) or weight of a structure. If the attribute is quantitative, it can be used in the generation of metrics (e.g., cost/tonne per day or cost/square meter). Attributes can also supplement understanding of the code of accounts; for example, the percentage of alloy piping could be captured in lieu of attempting to break out costs for alloy versus carbon steel piping. Key quantifies should also be captured for project. In summary, capture any attribute that supports meaningful queries and/or metrics depending on the expected uses of the data and metrics. Figure 6 is an example from the ICMS of attribute information (labeled “project” but primarily asset) for a water treatment plant.

April 1, 2021

Project Attributes	Values
Table 9: Waste Water Treatment Works (A facility for the cleaning and improvement of water that contains waste products, contaminants or pollutants to make it safe for discharge to land or water)	
Code	
Local functional classification standard	
• name of standard	
• code number of construction	
Works	
Functional type (descriptions of primary, secondary and tertiary treatment processes)	
Nature	new build | major adaptation
Environmental grade	
• grade and name of environmental certification	
• status	targeted | achieved | none
Principal design features	
• plant technology	
• number of processes	
• tank materials for each process	steel | concrete | other stated
• term of use	fixed | temporary
Project Complexity	
• standard of cleanliness of treated water (expressed in terms of significant parameters, e.g. Biological Oxygen Demand, Suspended Solids, etc.)	
Design life	(years)
Altitude	
• average height of site above or below sea level	(m | ft)
Dimensions	
• overall external diameter or length × width × height of each major structure	(m | ft)
Project Quantities	
Site area (area of land covered by permanent work, excluding temporary working areas outside the site)	(hectares | acres)
Functional units	
• capacity	(litres | gallons per day)

Figure 6: Example Asset Attributes for a Water Treatment Plant from ICMS [9]

2.3.5. Asset Requirements: Equipment

For the process industry, equipment is a major cost element of the plant investment. For example, perhaps the most recognized cost metric of interest in the process industry is the Lang factor; the ratio of total plant cost to the cost of equipment. Many process plant conceptual estimating techniques involve factoring from equipment costs at various levels of detail. Because of this, the capture of equipment costs and attributes of the equipment will typically be an important requirement of a process industry database; in essence, it is a database (tables) within the database. For each equipment type there will be a key code and attributes to capture (e.g., capacity, size, metallurgy, etc.). Equipment includes process and mechanical equipment such as vessels, reactors, compressor, pumps, motors, and so on, but also electrical equipment such as substations, transformers, and switchgear. The equipment data may be captured at the process unit level, but using equipment key codes, equipment data can be studied in its own right regardless of the plant or process it was used in. Figure 7 is an example of a data collection form to capture equipment attribute information (there are other forms shown in this reference) [18].

April 1, 2021

TECHNICAL FACT SHEET (one for each area)

BUDGET ID# AREA/UNIT TITLE:

TYPE OF WORK	
Enter approximate % of TOTAL CAPITAL $:	
1) NEW	
2) ADDITION or EXPANSION	
3) REVAMP or UPGRADE	
4) NON-PROCESS SITE/BLDG/UTIL	
5) DEMO/RELOCATE/REMOVE	
	0%

EXECUTION STRATEGY		
Enter approx. % of Construction Labor HOURs worked unter these conditions		
STANDARD	*Minimum OT, Non-shutdown*	
AGGRESSIVE	*Lots of OT, but not during shutdown*	
FAST-TRACK	*Engineering incomplete + Aggressive*	
SHUTDOWN	*Work actually during shutdown + OT*	
(OT = overtime - also covers other types of preminum paid work)		0%

PROCESS TECHNOLOGY	
Enter approximate % of TOTAL CAPITAL $:	
1) DEVELOPMENTAL or PILOT	
2) PROVEN BUT NEW TO COMPANY	
3) PROVEN, REPEAT	
	0%

RESOURCE STRATEGY		
Enter approximate % of appropriate CAPITAL $		
COMPANY	*COMPANY incl. temps and supplementary*	
ALLIANCE	*Formal alliance partners*	
CONTRACT	*Other than alliance*	
		0%

Note: The % values above MUST add to 100% - but rough guess of the breakdown is OK

PROCESS EQUIPMENT LIST (PRIME ACCOUNT "XXX")

List the major equipment items that predominantly explain the mat'l $ content of prime account AAA, BBB, and CCC

ID# (by est.)	TYPE	DESCRIPTION	QUAN #	CAPACITY	CAP. UNITS	TOT COST $
MAJOR ITEMS:						
Example:	*(e.g. Pumps)*	*(e.g. Zenith 20-103 pump assembly)*	*2*	*20*	*cc/rev*	*$ xx,xxx*
MINOR ITEMS (quantity and total cost only:						
					Subtotal:	$0

Note: The subtotal MUST = the sum of the AAA, BBB, and CCC material accounts in the project cost sheet!

INSTALLED PROCESS EQUIPMENT for which buy/fab was NOT covered in this project budget:

The Estimating Department can help you determine the $$$ value of these items!

Example:	*(e.g. Kettle)*	*(e.g. Existing Spare - Jacketed)*	*1*	*800*	*liters*	*$ xx,xxx*
TOTAL OF ALL EQUIPMENT ITEMS INSTALLED:						$0

ELECTRICAL & INSTRUMENT EQUIPMENT LIST (PRIME ACCOUNT "YYY")

List the major E&I items that predominantly explain the mat'l $ content of prime account DDD, EEE, and FFF

Primarily includes, Switchgear, Control or Power Panels, MCCs, PLCs, and Computer Hardware

ID# (by est.)	TYPE	DESCRIPTION	QUAN #	CAPACITY	CAP. UNITS	TOT COST $
Example:	*(e.g. Drive System)*	*(e.g. complete Reliance system)*	*1 lot*	*28*	*motors*	*$ xx,xxx*
						$0

Figure 7. Example Form to Capture Project and Equipment Attributes [18]

2.3.6. Project Requirements: Attributes (i.e., Practices and Risks)

Another attribute type to capture beyond asset or equipment attributes, which are mostly descriptive, are project system and execution practices. This includes data on execution strategies (e.g., fast-track), contract strategies (e.g., early contractor involvement), and practices in the area of team development, project control, procurement, safety and so on. These should be codified (e.g., code for each execution strategy type) or quantified (e.g., ratings of level of team development) as appropriate. Some attributes may be required while others are optional (e.g., for small projects) to avoid quality problems resulting from an overly onerous data collection process.

This data supports analytic research on how practices drive outcomes such as cost growth, schedule slip, safety metrics, quality metrics, operability metrics, sustainability metrics and so on. This research also supports tool development such as parametric models of systemic risks which requires that the systemic risk ratings (e.g., level of team development) be captured [19] (Chapter 11 - *Systemic Risks and the Parametric Model*). To support future risk identification and analysis, a good practice is to capture data on the top risk events and their impacts [19] Chapter 18 - *Closing the Loop*). Some may also capture lessons learned (for the purposes of this RP, this is assumed as a separate database/tool that would be linked to). There are examples in the literature of capturing extensive risk data [20] [21]. The stakeholders should be encouraged to envision how data about practices in their areas might be used to support practice improvement. Finally, it is useful to include some form of narrative which "tells the story" about the project. This helps contextualize the project data and inform the user.

2.4. Implementation

2.4.1. System Phasing and Migration Strategies

There are several reasons why a phased database implementation strategy should be considered. The first is that the number of projects at the start will be relatively small. The next is that it takes time to iron-out the database management process including defining a structure, developing and piloting data capture forms and mechanisms such as contract terms, and competency development in data handling and analytics. This can favor using spreadsheets at the start which staff are likely comfortable with, and likely already using to some extent. However, after the process is established, and the number of records increases, migrating to a relational database is recommended. Figure 8 is an example of database development phases that one company went through [22].

Q4 2008	2009	2010	2011
Requirement	**Proof of Concept Phase 1**	**Proof of Concept Phase 2**	**Implementation of PKMS**
Need for consistent capture and reporting of data	Project data is captured using excel spreadsheets developed by BDE Resulting reports prove useful but are very time-consuming to create	Software identification/purchase – development begins PKMS is approved as a PLGC deliverable	Customizations applied to meet Company's needs Migration of existing data & new data population Testing, Training & Communications Implementation & Ongoing support

Figure 8. Example of Database Development Phases Used by One Company [22]

2.4.2. Software Systems

Depending on requirements, there are two main options for data repositories and to host management (e.g., normalization) and application tools (e.g., estimate validation). Typical options for database repositories are spreadsheets and relational databases (NoSQL, a non-relational database, is also an option). Spreadsheets are usually created in-house. Relational databases can either be custom coded in-house or by a 3rd party or acquired as a commercial software system.

There is a third repository type that is the databases stored within BIM, estimating, scheduling and other design and planning software that retain past estimates and schedules. Scheduling software vendors may even offer metrics and diagnostic information derived from schedules of *all* its users stored in the cloud. However, for purposes of this RP, these external databases are considered sources of estimated data (not actuals) to which an historical database may be linked.

In addition to the database software, many will also desire visualization and/or business intelligence. These are not covered here, but these systems add graphical reporting and various analytical functions.

Note that phased implementations transitioning from spreadsheets to a relational database are common. The following are typical (i.e., not necessarily recommended) approaches and expectations and considerations for the options, including a strengths/weaknesses summary:

2.4.2.1. Spreadsheets

Spreadsheets are a basic option as a first step in a phased approach (i.e., spreadsheets are not recommended as a long-term approach). The main element is a standard spreadsheet template that project teams complete at the time of the estimate(s) and again as part of project closeout. The templates are checked by a trained staff member (tool administrator) for quality and stored in a central location, along with any backup files. The administrator maintains a master index spreadsheet with key field data to support process management and reporting and to locate applicable records for a specific analysis. Backup files are reviewed to check appropriateness of the record. At the time of an analysis (e.g., estimate validation), the user identifies comparison records from the index, then runs copies (originals are secured) of the selected templates through a normalization tool to adjust cost data to the analysis basis date and currency. The adjusted data is copied back to a place in the template designed for it.

For estimate reviews, a validation or benchmarking (re: Recommended Practice 110-20, *Cost Estimate Validation* [23]) spreadsheet tool calculates and applies the applicable metrics in comparison to the estimate. The user must edit validation tool formulae to draw non-outlier data from the respective templates that have been adjusted. The administrator will assist any user that is not trained. Basic graphics will be included in the validation tool. Business intelligence software is used to support advanced analytics and reporting. Special studies are done on an ad-hoc basis.

In summary, four spreadsheet tools are needed: the data capture templates, the master index, the normalization tool, and the validation tool. Other needs are the business intelligence software, and price indices for use in the normalization tool.

2.4.2.2. Relational Database-Custom Coded

This option is not commonly used now that viable commercial software is available, but there are situations where needs are unique enough to justify it. In that case, an in-house or 3rd party programmer(s) develops a relational database using software for that purpose (i.e., traditional coding may be minimal). Usually, data capture uses the spreadsheet templates with the database designed to upload quality checked spreadsheet data. The tool administrator generally is not the software expert/programmer, so support can be a challenge.

The database supports record search through queries of the data table key fields. Also, business intelligence software may be used to tap the database for data visualization purposes. The system may either be programmed to perform normalization (a programming challenge), or the records may be output to a normalization spreadsheet tool with adjusted data re-imported back to the system. The database is programmed to calculate metrics and to produce an estimate validation report and optimally graphics. Other reports such as time trend and performance reports could be programmed. However, like the normalization step, data could be exported for use in a separate validation spreadsheet or for other special studies. The administrator and user would work together to make this work.

In summary the tools needed include data capture spreadsheet templates (some inputs may come from system links), the database, and, if not programmed into the system, the normalization tool and the validation tool(s). Other needs are business intelligence software and price indices for use in the normalization tool.

2.4.2.3. Commercial Software

There are now viable project historical database systems with normalization capability on the market. All evolved from learnings by companies who first got involved developing custom-coded systems. These all have normalization (price indices must be acquired or developed), validation report and graphics capability. Data capture is often via spreadsheet templates which are quality checked by an administrator, and the database systems support spreadsheet import capability; however, data can also be directly input (but still quality checked). The tools all offer cloud capability, security and access control and so on. Some of the vendors offer other software products such as for estimating and project control; an advantage for companies that use that software as well. The main owner effort is working with the vendor in configuring the software. This primarily involves defining code structure, metrics, and special report design. Once set-up, it eases the use by multiple users and locations, and support is available from the vendor.

In summary the tools needed include data capture spreadsheet templates (some inputs may come from system links), the database software package/service, and, if not included in the software, the validation tool(s). Other needs are business intelligence software (for extension of built-in analytics) and price indices needed for normalization functionality.

2.4.2.4. System Strengths/Weaknesses

Table 4 describes strengths and weaknesses of the three software system options (note that any of these may be tapped into by visualization and/or business intelligence software).

Alternative	Strengths	Weaknesses
Spreadsheet	Entry point for phased implementation, Lower initial and annual cost, Flexibility, In-house.	Security, Integrity (errors), Support, User access and number, Scalability is very limited.
Custom Coded	Could have similar or greater capability as commercial but depends on quality of the implementation. Typically, higher initial cost than other options, but lower annual cost than commercial if changes are limited over time.	Software projects are risky and may have quality issues requiring continual updates and maintenance. Lack of experience in developing project history systems. Annual costs may be low, but system may become outdated and unsupportable.
Commercial	Scalability, Multiple users, Ease of access, Security, Integrity (less error), Support, Less staff effort.	Contracts, Less flexible, Higher annual cost but lower initial cost. Customizing can be costly (iron out the process first).

Table 4. Strengths and Weakness of the Project Historical Database Software System Options

2.4.3. Alignment with Building Information Modeling Systems

Building information modeling is used for the design or engineering of most major construction projects in the building and infrastructure industries (i.e., architecture, engineering, and construction (AEC)) and increasingly in the process plant world (i.e., engineering, procurement, and construction (EPC)). BIM extends digital modeling (e.g., computer-aided design (CAD)) to include cost and schedule data related to the asset. The following are the typical BIM dimensions:

- 2D – Computer Aided Design (CAD) – software used to produce drawings.
- 3D – Geometry - Three geographical dimensions (x, y, z) (the model) – object-oriented design.
- 4D – Time - duration, timeline and scheduling – phase and sequence of construction simulation.
- 5D – Money - cost estimation, budget analysis – estimating construction/capital.
- 6D – Sustainability - self-sustainable and energy efficient – energy management.
- 7D – Facility Management – information, operations and life-cycle maintenance.

The project historical database should provide linkages to the BIM databases for relevant historical information. Importing all of the BIM information into a project historical database is not practical or efficient. Generally, a historical project database would provide linkage to the digital modeling system (BIM or CAD) for access to reference final design drawings, specifications, and 3D models. For BIM, this linkage would extend to the final 4D schedule simulation and 5D estimate and basis of estimate (i.e., the core cost and schedule data for planning use). The linkage may not be direct, but through software that serves as a bridge from BIM to the database and/or directly to estimating and scheduling software. Finally, there should be linkage for access to final 6D energy management plans and 7D operations manuals and maintenance schedules for reference.

Some organizations may want to link to interim or phased deliverables as part of the historical project database to understand design and plan evolution. Note that the BIM data is planned data, and not final actual cost and schedule which must be obtained from other cost and schedule systems (e.g., ERP, project control, etc.).

A team integrating with BIM is likely better positioned to implement a project historical database. BIM is driving the digitalization of products and processes via standardized designs and data structures. BIM modeling will drive earlier and more data-driven decision making, online and end-to-end software platforms, advanced logistics management, more collaboration and other process improvements.

2.4.4. Alignment with Enterprise Resource Planning and Project Management/Control Systems

A key source for actual data is (or should be) the company ERP system. ERP systems (or equivalent) capture capital budget data and actual cost/resource information. For contractors, or for self-performed work, this may be a job-cost system or equivalent. There may be different systems for different regions or businesses. Finance/accounting departments usually have primary responsibility for these systems, and system design tends to focus primarily on their needs. However, as the words *enterprise* and *resource* indicate, the systems may encompass contracting, procurement or material management, human resources, and in some cases project management modules.

Procurement systems, sometimes embedded in ERP, are potentially valuable direct data sources if they are set up with material and equipment type key codes; and capture key attribute data. If these systems do not have these, material and equipment data may have to be manually extracted from equipment data sheets, quotes and/or purchase orders.

As mentioned, the ERP system may include a module for project management. More often, an owner company will have a system specifically for project control purposes that itself may be integrated with the ERP system for the capture of expenditure information. These project systems are likely to have data structured that is more usable for an historical database than from accounting systems. However, owners often manage projects by contract and not by discipline, and the WBS used will often not align with the asset type structure required for historical data and metrics (e.g., by process unit).

In summary, the content, structure and integration of ERP, procurement, project management/control, and equivalent databases vary widely. Therefore, each project database implementation effort needs to assess the quality of data and information available and make allowance for the interface required for data capture. This may range from filling out project and equipment data capture forms for actuals (typically spreadsheets). It often involves manually translating the ERP and other system data into a usable structure (including attributes) to direct system interfaces using structured files published from the ERP or other system (including quality checks).

Senior management's attention to data integration is increasing; therefore, it is likely that one will find the company in the midst of some sort of major process and system improvement effort. If so, it is important that historical database needs and requirements are considered in those efforts.

2.4.5. Organization/Roles/Responsibilities/Skills and Knowledge

Planning, implementing, and managing a database starts with having clear objectives. These should have buy-in from senior management, including from those responsible for evaluating and making capital project investments or bidding decisions as well as those responsible for the capital project management system. Senior management determines the leadership of implementation projects and administration of ongoing database management. Failure of organizational improvement projects (such as database implementation) can often be tied back to an unsupportive culture and/or lack of sustained executive leadership.

<u>2.4.5.1. Implementation Project Team</u>

While elements of a database may grow organically from various estimating improvement efforts over time (e.g., ad-hoc maturity), this RP assumes that at some point a formal project (or at least a budget request and resource commitments) is needed to reach managed or robust maturity. In that case, a project manager or team leader is assigned by senior management and a project team is identified. In addition to the PM or leader, the person(s) assigned responsibility for ongoing database management helps lead the team.

April 1, 2021

A development team benefits from having a number of competencies represented. In particular, people with estimating and other relevant domain experience (e.g., scheduling, capital planning, etc.) are needed. For owners, conceptual estimating experience is desired. Optimally, this person(s) should also have the broad range of skills and knowledge such as reflected in RP 11R-88, *Required Skills and Knowledge of Cost Engineering* [24]. In addition, a person(s) with analytics experience is desirable including at least basic skills and knowledge of statistical analysis. If the strategy is to implement a relational database, in-house or commercial, there will be a need for IT representation to address the software aspects of the project. If commercial software is to be used, once it is identified, the vendor should make available a subject matter expert (SME) to support the team. If the system strategy is spreadsheet based, the team will require spreadsheet application experience. In any case, an historical project database SME can help guide the development effort. This is the core team.

The core team can be supplemented with people experienced in various data use applications (e.g., from business planning, planning and scheduling, risk management, procurement, etc.). To ensure input from the user base, a working group of business and regional representatives may be formed, particular at the start as requirements and the basic process are defined. Sub-team(s) may also be formed for developing applications such as an estimate validation tool.

2.4.5.2. Database Operation and Maintenance

Senior management decides on the organization(s) charged with administration and operation of the database. A centralized organization will help ensure standardization and an effective shared application. The central organization also needs to be an unbiased broker of information to assure the needs of various business units and regions are balanced and met. A database manager or administrator is assigned by that organization. At the start, that person is part of the implementation team. Depending on the software strategy and scale of the database scope, the manager/administrator may need a functional group to manage operations and maintenance. For simple databases with limited uses, the manager/administrator may be a senior estimator managing the operations as a part-time effort, but this often evolves to a larger group. Transactional steps (e.g., data capture at project closeout) should be managed by the respective business processes (e.g., project control) in alignment with well-defined database procedures.

For a larger system with multiple uses, the operations group will typically have several full-time and/or part-time staff. Some key roles include:

- Facilitating data capture, cleaning, and normalization.
- Conducting and supporting analyses of various types for various data users.
- Coordination with IT and/or vendor software support.
- Gather learnings for potential improvements and leading system modification and maintenance projects.

2.5. Data Capture

Before a system is operational, it must be populated with data. Data must be captured, cleaned, and entered into the system. There are two phases or modes of capture, cleaning, and entry: data backfill and ongoing data production.

2.5.1. Backfill (Legacy Data) Phase

Until a system achieves a robust volume and variety of applicable, appropriately detailed, cleaned, and normalized data and metrics, it will not satisfactorily achieve its purposes. This requires initial database backfilling with cleaned/normalized past records at the start of database operation. As a rule-of-thumb, it takes greater than five quality, applicable observations in a comparison dataset before a metric is meaningful (but more is desirable). The backfill phase begins immediately upon project inception by doing an inventory of past projects and retaining files for the candidate projects. The following criteria characterize good data backfill candidates:

- Quality: Records are likely to have well organized and complete data.
- Applicability: Similar in scope to those in the upcoming project portfolio.
- Recency: The prior criteria are more important, but more recent data is better than old.

There is a trade-off between database uses and data backfill requirements. For example, if one has scores of good records of past bids in a defined structure, and limits use of the database to benchmarking bids, then data backfill effort is minimized to tasks such as attribute definition. However, if additional database uses are planned, and the bid records do not contain the needed data (e.g., schedule, actual costs, etc.), then data backfill effort is increased.

The selection of project data samples should reflect a more or less random distribution of performance. Once identified, all applicable files should be gathered and retained for data processing. This includes not only records like closeout cost analyses, but files such as change logs and risk registers that will provide information useful for data allocation and other learnings.

Once the structure is decided, and a data capture cost form is developed, the data cleaning and form completion (allocation) process can begin. Data cleaning and allocation is always required; rarely can unprocessed data be used in its entirety. If the code of accounts has been changed over time, legacy project data must be aligned with the current code of accounts. Thus, allocation of costs must be considered for both code of accounts and asset breakdown structures. Another challenge is allocating/segregating indirect costs that may be embedded in a cost item (the basis of estimate should explain what costs may have been embedded). Costs need to be allocated by asset type which may not match the COA used for project control. This is not a trivial task; it requires estimating competency with the ability to glean evidence from the project closeout records to support reasonable allocation. For example, if actual concrete hours from a contractor's records cannot be found, the estimated hours can be used as a starting point, and reasonable assumptions made as to later changes in hours or cost by looking at the change management records. The actual cost data range (as well as indirects) may be within the ranges of Class 4 accuracy, which puts a premium on large sample sizes to achieve statistical significance.[2]

2.5.2. Production Phase

Once the system has been tested using the backfill data, the data collection can begin from projects in progress. By this point the project teams and contractors should be completing raw data capture forms or equivalent at each decision gate as part of their responsibility. Some may wish to also capture data snapshots and perform benchmarking during the course of the project (e.g., periodic or at key milestones). At project close-out, the project control lead will provide the actual data capture form or equivalent.

Expect that new raw data will need cleaning, as well as some allocation, and quality control. If structures and systems are truly integrated (e.g., estimate data published from a BIM model), some of these steps can be streamlined or

[2] This backfill effort, and its learning curve, will require significant resources and time. One reference says the effort was "painful" and took over six months with a small team to capture just seven projects. In some cases, the project team (i.e., culture) was an impediment; i.e., arguing over how "their" data was being re-allocated (a challenge to their process and integrity). Others will insist on linking to and data dumping from other accounting and/or control systems [8]. This points to the need for good stakeholder management from the start and starting early. Once the backfill is done, the process should be ironed out and all parties on the same page.

semi-automated. Organizations may need to change their work flow (including close-out procedures) in order to capture historical costs.

2.5.3. Data Capture Forms or Publishing Formats

Once a structure is decided, and end uses and data requirements defined, then forms for data capture are developed. If the capture is via an integrated system (e.g., from BIM or ERP) then this step is about defining and setting up the data publishing format, not a form. There may be separate forms/formats for different asset types (i.e., different attributes). There may also be different forms/formats for estimates and for actuals. The backfill process should resolve problems with the forms or formats. Typically, data is collected by spreadsheet since most database software allows easy import of spreadsheet data into the system. During the production phase, the project control team will be responsible for form completion, or equivalent, and validated by the project manager or designate.

2.5.4. Data Cleaning

Data cleaning is the process of converting raw data into a form that meets system content and quality requirements. This does not include data normalization. Data cleaning starts with defining acceptance criteria. It is best to capture all projects regardless of performance. However, one should note the impact of extraordinary risk events. Data cleaning consists of correcting accounting and other errors in the raw data, and allocation to database structure requirements. Any changes to the dataset should be reconciled and approved by the appropriate authority.

2.5.5. Project Closeout/Backup Documents

Historical cost data should include supporting documents, records and files to assist with cost data allocation, analyses, learning, and dispute resolution support. Typical examples of supporting documentation are: the bases of estimates and schedules, schedule of values, change logs, risk registers, etc. Much of this may already be obtained as part of the company's project closeout process and managed by a document control function. If document control is of good quality and readily accessible, there is no need to create redundant systems. Therefore, project closeout is a key function of the data collection process. Figure 9 shows an example of a list of documents retained to support an historical database process. This list will vary for every company depending on its requirements.

i.	Project Summary	
	a.	Basis Project Fact Sheet
	b.	Project Narrative – 1 to 3 page write-up
	c.	Cost Breakdown Structure – Summary Report
ii.	Scope Documents	
	a.	Requirements Document
	b.	Basis of Design
	c.	Project Organization Chart
	d.	Preliminary and Final Funding Documents
	e.	Key Quantities or Metrics
iii.	Schedule	
	a.	Basis of Schedule
	b.	Final Master Schedule
	c.	Special/Unique Control Level or Detail Schedule
	d.	Engineering and Design Progress Report
	e.	Construction and Start-Up Schedule
iv.	Project Estimate	
	a.	Project Estimate
	b.	Basis of Estimate
	c.	Project Breakdown Structure Summary
v.	Cost and Performance	
	a.	Summary Cost Reports
	b.	Final Cost Control Reports
	c.	Final Code of Accounts Report
	d.	Expenditure Plan and Cash Flow Curve (s)
vi.	Contracts	
	a.	Contract (s)
	b.	Contract Bid Summaries
	c.	Bid Form
	d.	Subcontract (s)
	e.	Unit Price and Other Detail Cost Submittals
vii.	Technical or Design Information	
	a.	Plot or Site Plan
	b.	Block Flow Diagrams
	c.	Process and Instrumentation Drawings
	d.	Drawing List
	e.	Links or Attachments – Drawings, Specifications, and Other Documents

Figure 9. Example Database Backup Document List (adapted from [7]**)**

2.6. Data Processing, Analytics/Metrics and Reporting

2.6.1. Normalization

Typical uses of data, such as metrics for estimate development or validation, call for a like-for-like comparison. Normalization is the process of adjusting data for time, currency, and location:

- Time: this adjusts for price escalation. The topic is addressed in RP 58R-10, *Escalation Estimating Principles and Methods Using Indices* [25].

April 1, 2021

- Currency: this adjusts for differences in currencies (exchange rate) between the project and the reference currency. Exchange rates change over time, and are driven by similar economic conditions as escalation, so the adjustment processes are related. An example of this can be seen in [26].
- Location: this adjusts for differences in locally sourced labor rates and material costs between the project and the reference location. For example, unit masonry is usually locally sourced, and local construction labor will have local hourly rates and productivity. The adjustment is for *prevailing* regional productivity, not for project-specific performance differences. Costs for services, materials and equipment that are globally sourced are typically not adjusted. This may also involve adjusting for differences in regulatory codes and specifications. The topic is addressed in RP 28R-03, *Developing Location Factors by Factoring – As Applied in Architecture, Engineering, Procurement, and Construction* [27].

These are the most common normalizations, but others may include: measurement system (metric vs. imperial), complexity adjustments, or others required to meet stakeholder requirements.

It is generally not necessary to perform normalization until the time of data application. The data that resides in the database is "as captured" and the key fields of date(s), currency(s) and location record the basis of that data. These key fields are used in the normalization calculations. For example, if an estimate validation is to be performed using the metric of costs/tonne of steel, the costs would be adjusted for escalation and currency when the metric ratio is calculated. Raw data is always retained but normalized data is used for analysis.

It should be noted that normalization does not adjust for differences in scope. Data needs to be captured by meaningful asset or scope types (e.g., for process units or areas in addition to the overall plant or facility). Getting a like-for-like comparison on scope is a function of creating the best comparison set using the asset and project attributes in queries of the database.

For example, to describe the normalization steps, consider the need to validate a cost estimate of material cost/cubic-meters of concrete that has a basis date of 2020, currency of Euros, and located in Rotterdam. In this database consider that there is a record of *actual* concrete material cost for a project completed in 2017, in US dollars, and located in Houston.

Normalization will adjust the database record to the estimate basis. The following steps apply:

1) Time: As an actual cost, estimate the mid-point of spending on concrete material (e.g., if concrete work was done in 2016, that is the date that applies, not 2017). Escalate the 2016 US concrete material cost to 2020 using US price indices (price indices of the country/region where the project is located).
2) Currency: Adjust the 2020 US concrete material cost to Euros using the 2020 exchange rate.
3) Location: Being a locally-sourced material, adjust the Houston concrete cost to a Rotterdam basis using material unit cost data from reliable sources for each location (either from published cost data books or factors developed in-house through analysis of past company data)
    a. If the metric was labor hours/cubic-meters, time and currency would not apply, but the location adjustment would be for relative location productivity. This would be based on published unit labor norms from books or factors developed in-house through analysis of past company data.
    b. If the metric was labor cost/cubic-meters, it would be adjusted for time and currency per steps 1 and 2 with an additional step of adjusting for relative all-in labor rates, based on published or in-house labor rate data and factors.

As can be seen, the location adjustment requires reference data of unit cost and labor norms by region. Companies with many international projects should acquire and maintain a library of regional reference data. The applicability and/or quality of published location data may be suspect [28]; therefore, once the database is well populated, conduct location studies to develop standard adjustment factors. It is recommended to not include items such as taxes and duties in the normalization; these are captured in separate cost accounts and not embedded in the item costs.

When using actual data to compare to an estimate, another consideration is that the actual costs are expected to be greater than a *base* estimate, because actual cost includes the impacts of risk (i.e., the consumption of contingency). This must be considered in the final analysis. For example, if the estimate mean metric is marginally below the actual mean in the estimate validation, that is expected; do not increase the estimate because in the end, contingency spending will make up the difference.

In some cases, the order of the normalization steps 1 and 2 listed above could be reversed. In the example, currency could first be converted from dollars to Euros, and the result escalated using the price indices for Rotterdam. The result will be different if the price trends vary by location, and/or currency exchange rates change markedly over time. Assume that allocation and adjustments render the metrics equivalent to AACE Class 4 quality (i.e., statistical noise may be introduced by calculation methods, putting a premium on having large comparison datasets). The order of calculations should remain consistent for the historical database.

While the steps of the process above may seem relatively straight-forward, there are many nuances, particularly with the issue of proper price index selection/development for escalation. As is discussed in RP 58R-10, *Escalation Estimating Principles and Methods Using Indices*, published price indices often require adjustment before application in the normalization process. Often there are no available indices for the prices of services that are bid; and they are highly sensitive to market conditions. RP 58R defines a method to adjust the published indices for the market.

The choice of software application for the database will often impact the normalization process. For spreadsheets, normalization is usually performed in a separate spreadsheet-based normalization tool. Each time an analysis is prepared (e.g., estimate validation), it is necessary to manually convert each comparison project record to the desired basis, project-by-project. For programmed relational databases, the normalization functionality (including updated price indices, exchange rates and location factors) is usually built into the software, which makes the analyses more efficient.

If normalization is performed properly, with accurate price indices and factors, data records can be maintained and used for many years. For example, one benchmarking firm has been maintaining a database for over 30 years [29]. Many metrics for engineering and construction are surprisingly steady even over the period of decades.

2.6.2. Analytics/Metrics and Data/Metrics Export

The database itself is primarily used to capture and store structured data that can be easily queried, organized, and reported. It should support required normalization. The database should support the generation of information in the form of metrics, which are usually ratios of data on one resource to another (e.g., duration, costs, quantity).

A key application of historical data is to aid in research (i.e., studies that correlate practice data to outcome metrics). This supports continuous process and practice improvement (e.g., the application of machine learning). If these functions are external to the historical database system, they require that the historical database supports the export of raw or normalized data, as well as metrics, in a format that other applications can use (preferably through direct application programming interfaces (APIs), but commonly as spreadsheet tables).

The database may also support other various applications and functions. A common use of the historical database is to support estimate validation (of cost or duration), which compares an estimate's metrics to those of a comparison dataset. RP 110-20, *Cost Estimate Validation,* documents this usage and also provides more discussion of appropriate metrics to use [23]. Other functions are usually external to the historical database such as cost estimate preparation and parametric model building for cost, duration, or risk.

2.6.3. Reporting/Communication/Graphics

The database is primarily used to capture and store structured data that can be easily queried, organized and reported. This data may also be used in other applications. Many historical systems are designed to support basic reporting of metric *descriptive* statistics and time trends. Descriptive statistics are means, median, ranges, etc., which describe the distribution of data. These statistics may support understanding of a particular project (e.g., project cost account breakout versus breakouts of a comparison set) and are often used by management to assess and benchmark project system performance (i.e., comparing metrics by business unit or region) and improvement (metric changes over time). *Inferential* statistics (e.g., regression) is usually performed externally using specialized statistical software applied to the exported data.

Descriptive statistics are usually presented graphically to aid communication. These include the usual bar and pie charts, scatter diagrams, trend charts, etc. Increasingly, visualization and business intelligence software facilitates this capability by querying the database system. Figure 10 provides an example of a metric comparison ($/diameter-inch-mile) for four observations from a database system [22].

April 1, 2021

Cost Category	Project 1	Project 2	Project 3	Project 4
01. Pipeline Material	$21,462	$24,431	$15,312	$25,051
02. General Construction	$54,397	$58,021	$42,543	$38,968
05. Land	$4,442	$4,964	$3,140	$5,104
06. Project Control	$8,352	$9,026	$7,707	$8,940
07. Contingency	$8,865	$0	$8,865	$7,806
08. Project Others	$5,291	$5,359	$5,325	$55
09. Total Installed Cost, excluding AFUDC[3] and Insurance (01-07)	***$97,518***	***$96,442***	***$77,567***	***$85,869***
10. Total Installed Cost, including AFUDC and Insurance (01-08)	***$102,809***	***$101,801***	***$82,892***	***$85,924***

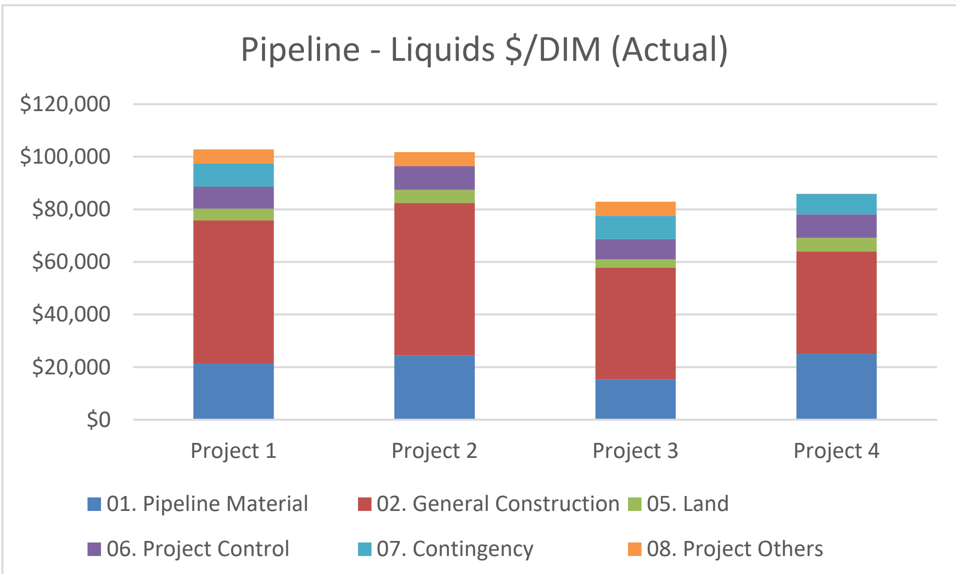

Figure 10. Example of a Metric Comparison Report from a Database System (adapted from [22]**)**

2.7. Data and System Quality, Training, and Maintenance

2.7.1. Data and System Quality

Quality control is required for any process or system. As part of the implementation, a test plan should be developed to ensure data integrity and consistency. This includes testing of inputs, normalization calculations, metrics calculations and reporting. During testing, anomalies of the data capture and cleaning process may be noted and fixed. During operation any quality issues should be recorded and corrections made as needed. There may be periodic, or milestone tests performed (such as each time revised price indices are loaded, or a new company entity or region is brought into the process).

[3] Allowance for funds used during construction (AFUDC). The net cost of borrowed funds and a reasonable rate on other funds used during the period of construction.

2.7.2. Training

Once a system is in use, the appropriate stakeholders need to be trained in data capture, cleaning, software maintenance, and in the use of the application. Training may be required in order to produce periodic reports on project cost and performance trends, which should include training about what the reported information conveys. Organizational planning should assure that there is continuity in the skills and knowledge of the system. As the database evolves and expands, new uses will be found and there will be new stakeholders to train.

2.7.3. Maintenance

Once a system is in operation, there will ongoing and periodic updates to the software to keep current with evolving information technology and market developments. If commercial software is used, this may be accomplished largely by vendor support, but at times it will become necessary for the customer to get directly involved with the process. As the volume and diversity of data increases and company needs evolve, more uses of data and ways to analyze and report them may be required to be developed.

A potential major disruption is any changes to code of account structures and systems used for project control. Generally, changes to the database should be avoided. However, it may become necessary that some elements of data need to be recast into a modified structure.

Another maintenance issue is data that becomes obsolete; these may be deleted from the database as they are identified. However, effective normalization should typically allow old data to remain relevant for many years. There is a tradeoff between the benefits gained in statistical significance by having larger datasets and the quality issues that arise with using older data. Many metrics will be found to be surprisingly stable over time (unfortunately, this may reflect a lack of improvement in design or execution performance). In any case, keeping older data (even though it may be obsolete or irrelevant) to study time trends can be informative.

2.8. Conclusion

This recommended practice defines the basic elements of and provides broad guidelines for evaluating, developing and maintaining project historical data management systems (i.e., a database). The goal is to help organizations in planning a new historical database implementation or in accomplishing a significant advancement in database maturity. This RP describe special considerations for defining database asset and implementation project scope and steps including its life cycle operation and management (i.e., to provide a basis or framework for planning). Guidelines for using data from the database such as estimate validation, estimating databases, parametric modeling and so on are addressed in other AACE RPs.

REFERENCES

[1] H. L. Stephenson, Ed., Total Cost Management Framework: An Integrated Approach to Portfolio, Program and Project Management, 2nd ed., Morgantown, WV: AACE International, Latest revision.

[2] AACE International, Recommended Practice No. 10S-90, Cost Engineering Terminology, Morgantown, WV: AACE International, Latest revision.

[3] I. Bromberg, G. E. Brooks, W. J. Dodge, E. Eaves, A. Lothrop, P. Mayo, D. Rogers and R. E. Templeton, "Essentials of Cost Control for Capital Plant Expenditures," in *AACE Transactions*, Morgantown, 1960.

[4] G. Samid, "Cost Database: How To Build Or Buy The One You Need," *Cost Engineering,* vol. 26, no. 02, 1984.

April 1, 2021

[5] J. F. Enrico, "(SK.3) Project Control - Data, Information, and Computers," in *AACE Transactions*, Morgantown, 1991.

[6] National Research Council, "Report From the 1986 Workshop on Integrated Data Base Development for the Building Industry," The National Academies Press, Washington, DC, 1987.

[7] J. K. Hollmann, "(MAN.2) Project History - Closing the Loop," in *AACE International Transactions*, Morgantown, 1995.

[8] C. Figueiredo and R. S. Philipenko, "(OWN.09) Developing and Implementing a Project Knowledge Management System (PKMS)," in *AACE International Transactions*, Morgantown, WV, 2010.

[9] International Construction Measurement Standards Coalition (ICMSC), "ICMS: Global Consistency in Presenting Construction and Other Life Cycle Costs, 2nd ed.," International Construction Measurement Standards Coalition (ICMSC), London, United Kingdom, 2019.

[10] W. G. Hombach and P. C. Sanford, "(IT.1274) Building a Cross-Cutting Historical Cost Database for Mega-Projects," *AACE International Transactions,* 2013.

[11] Construction Specifications Institute (CSI) and Construction Specifications Canada (CSC), Masterformat, 2020 ed., Alexandria, VA: CSI, 2020.

[12] Construction Specifications Institute (CSI), *Uniformat,* Alexandria, VA: CSI, 2010 ed..

[13] International Standards Organization (ISO), ISO 19008:2016, Standard cost coding system for oil and gas production and processing facilities (formerly Norsok Standard Z-014), 2016.

[14] AACE International, Recommended Practice No. 21R-98, Project Code of Accounts - As Applied in Engineering, Procurement, and Construction, Morgantown, WV: AACE International, Latest revision.

[15] AACE International, Recommended Practice No. 103R-19, Project Code of Accounts - As Applied in the Mining and Mineral Processing Industries to Facilitate Benchmarking, Morgantown, WV: AACE International, Latest revision.

[16] P. C. Sanford, M. A. Moe, T. J. Brennan and W. G. Hombach, "Environmental Cost Analysis System Data Formulation - 10288," in *WMSYM*, Phoenix, AZ, 2010.

[17] Construction Specifications Institute (CSI), OmniClass: A Strategy for Classifying the Built Environment - Introduction and User's Guide, 2.1 ed., Alexandria, VA: Construction Specifications Institute (CSI), 2019.

[18] D. B. Roggenkamp, "Use and Implementation of a Project Historical Database System," *Cost Engineering,* vol. 45, no. 10, 2003.

[19] J. K. Hollmann, Project Risk Quantification, Sugarland, TX: Probabilistic Publishing, 2016.

[20] B. Skokan, D. Melamed, R. Lehman and J. Lefman, "(RISK.1284) Tailored Project Risk Management Assessment Tool," in *AACE International Transactions*, Morgantown, WV, 2013.

[21] B. Skokan, C Liu, R. Lehman, J. Lefman and D. Melamed, "(RISK.2473) Lessons Learned in Completing a Tailored Project Risk Management Assessment Tool,," in *AACE International Transactions*, Morgantown, WV, 2017.

[22] F. C. Figueiredo and R. S. Philipenko, "(OWN.772) Taking a Project Knowledge Management System (PKMS) to the Next Level," in *AACE International Transactions*, Morgantown, 2011.

[23] AACE International, Recommended Practice No. 110R-20, Cost Estimate Validation, Morgantown, WV: AACE International, Latest revision.

[24] AACE International, Recommended Practice No. 11R-88, Required Skills and Knowledge of Cost Engineering, Morgantown, WV: AACE International, Latest revision.

[25] AACE International, Recommended Practice No. 58R-10, Escalation Estimating Principles and Methods Using Indices, Morgantown, WV: AACE International, Latest revision.

[26] P. A. Cronje, "(EVM-3265) Earned Value Challenges in Projects With Different Project and Functional Currencies," in *AACE International Transactions*, Morgantown, WV, 2019.

[27] AACE International, Recommended Practice No. 28R-03, Developing Location Factors by Factoring, Morgantown, WV: AACE International, Latest revision.

[28] B. Bradley, "(EST.05) What is the U.S. Gulf Coast Baseline when Estimating Site Productivity?," in *AACE International Transactions*, Morgantown, WV, 2008.

[29] F. Biery and M. Stewart, "(EST.1619) Benchmarking Mining and Minerals Processing Projects," in *AACE International Transactions*, Morgantown, WV, 2014.

[30] P. R. Bredehoeft and H. L. Stephenson, "(TCMA-3459) The Unit Price Process for Estimate Data Development and Benchmarking," in *AACE International Transactions*, Morgantown, WV, 2020.

[31] H. L. Stephenson and P. R. Bredehoeft, "(TCMA-3502) Benchmarking for Competitive Advantage," in *AACE International Transactions*, Morgantown, WV, 2020.

[32] W. E. Kraus, "(EST.1579) A DIY (Do It Yourself) Unit Cost Database," in *AACE International Transactions*, Morgantown, WV, 2014.

[33] S.-H. Lee, M. T. Reiland and J. D. Whiteside, "(DEV.1634) Benchmarking and a Methodology for Finding Causality," in *AACE International Transactions*, Morgantown, WV, 2014.

[34] M. Hu, J. G. Krishnan and Y. Wang, "(TCM-2232) Alberta Transmission Project Cost Benchmarking," in *AACE International Transactions*, Morgantown, WV, 2016.

[35] T. W. Pickett and B. G. Elliott, "(EST.02) Transforming Historical Project Data into Useful Information," in *AACE International Transactions*, Morgantown, WV, 2007.

[36] B. G. Elliott, "(EST.02) Project Historical Databases for the Canadian Oilsands," in *AACE International Transactions*, Morgantown, WV, 2005.

CONTRIBUTORS

Disclaimer: The content provided by the contributors to this recommended practice is their own and does not necessarily reflect that of their employers, unless otherwise stated.

Peter R. Bredehoeft, Jr. CEP FAACE (Primary Contributor)
John K. Hollmann, PE CCP CEP DRMP FAACE Hon. Life (Primary Contributor)
Fred J. Wellman (Primary Contributor)
Christopher W. Carson, CEP DRMP PSP FAACE
Larry R. Dysert, CCP CEP DRMP FAACE Hon. Life
Bruce G. Elliott, CCP FAACE
William G. Hombach
Dave Kyle, CCP CEP
Michael Lesnie
Arvid Markus
Dan Melamed, CCP EVP FAACE
Nick Papadopoulos
Todd W. Pickett, CCP CEP FAACE
Avi Sharma
H. Lance Stephenson, CCP FAACE
Subhash Tuladhar

John K. Hollmann, PE CCP CEP DRMP FAACE Hon. Life
Validation Estimating LLC
jhollmann@validest.com

Peter R. Bredehoeft, Jr. CEP FAACE
HDR Inc.
peter.bredehoeft@hdrinc.com