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
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

PROJECT HISTORICAL DATABASE DEVELOPMENT

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

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 (TCM)* [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.

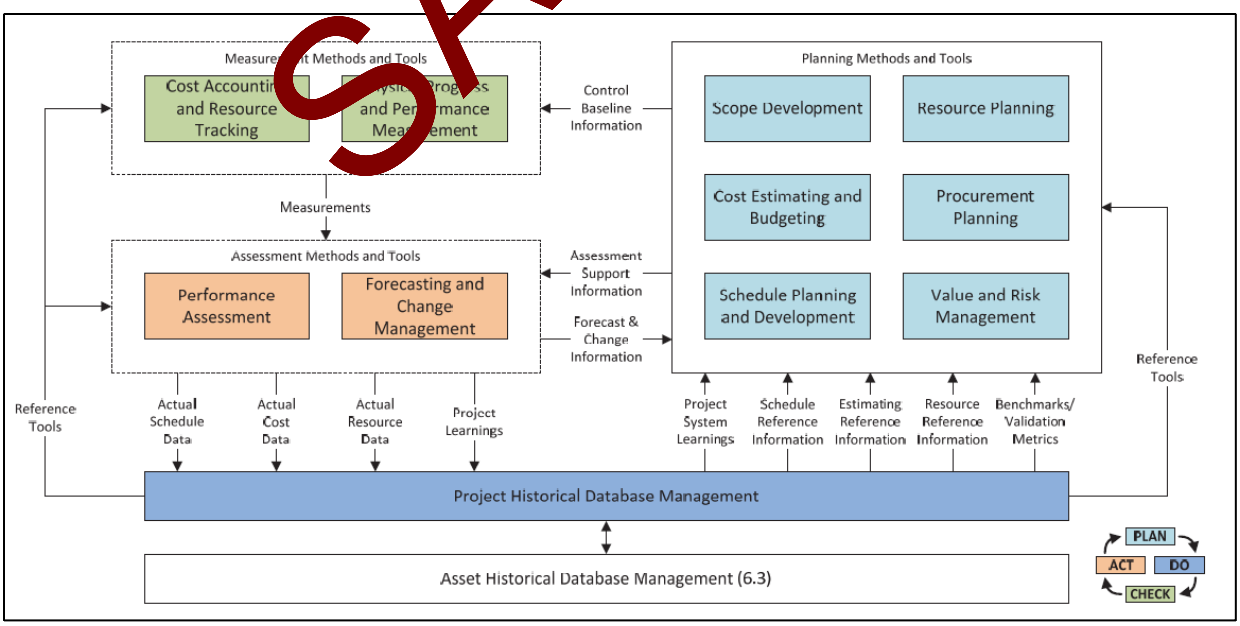

Figure 1: TCM Project Control Information Flow Supported by a Database [1]

April 1, 2021

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*

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 valuation 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 comp set in the 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 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".