Available online at www.sciencedirect.com

ScienceDirect

International Journal of Project Management 33 (2015) 697 – 710

International Journal of Project Management

www.elsevier.com/locate/ijproman

Construction and evaluation framework for a real-life project database

Jordy Batselier [a], Mario Vanhoucke [a,b,c,*]

[a] Faculty of Economics and Business Administration, Ghent University, Tweekerkenstraat 2, 9000 Ghent, Belgium
[b] Technology and Operations Management Area, Vlerick Business School, Reep 1, 9000 Ghent, Belgium
[c] Department of Management Science and Innovation, University College London, Gower Street, London WC1E 6BT, United Kingdom

Received 16 May 2014; received in revised form 18 August 2014; accepted 9 September 2014
Available online 6 October 2014

Abstract

In this paper, a real-life project database is created, outranking the existing empirical databases from project management literature in both size and diversity. To ensure the quality of the added project data, a database construction and evaluation framework based on the so-called project cards is developed. These project cards incorporate the concepts of dynamic scheduling and introduce two novel evaluation measures for the authenticity of project data. Furthermore, an overview of the constructed database leads to statements on the difference between planned and actual project performance and on the earned value management (EVM) forecasting accuracy. Moreover, the database is publicly available and can thus become the basis for many future studies related to project management, of which a few are suggested in this paper. To further support these studies, the database will continuously be extended utilizing the project cards. Furthermore, the project cards can also serve didactical purposes.

Keywords: Project management; Dynamic scheduling; Earned value management; Empirical database; Database construction framework

1. Introduction

The concept of *dynamic scheduling* was introduced by Uyttewaal (2005) and further extended by Vanhoucke (2012b) and incorporates the three dimensions of project management: baseline schedule, risk analysis, and project control. The *baseline schedule* expresses the planned course of a project and acts as a central point of reference for the two other dimensions. The search for the best possible baseline schedule under certain resource restrictions is known as the resource-constrained project scheduling problem (RCPSP). During the *risk analysis* phase, Monte Carlo simulations are used to generate activity durations and costs that deviate from their baseline values in order to assess the impact of these deviations on the time and cost objectives of the project. This technique is also called schedule risk analysis (SRA) and was first introduced by Hulett (1996). *Project control* then consists of the monitoring of the performance of a project in progress and comparing it to the baseline schedule with the aim of triggering corrective actions when the project's objectives are jeopardized. Earned value management (EVM), with the incorporation of the earned schedule (ES) concept, is the preferred technique for performing project control. This technique will more elaborately be discussed in Section 2.1.4.

In the last few years, there has been extensive high-quality research on each of the three dimensions of dynamic scheduling, which proves their relevance to modern-day project management. Concerning baseline scheduling, Hartmann and Briskorn (2010) provided an overview and classification of the most important extensions of the RCPSP. More recently, several project scheduling researchers focused on one particular extension of the RCPSP, namely the multi-mode RCPSP (MRCPSP), in which multiple execution modes are available for each activity in the project (Coelho and Vanhoucke, 2011; Deblaere et al., 2011; Elloumi and Fortemps, 2010; Van Peteghem and Vanhoucke, 2010, 2014; Wang and Fang, 2011, 2012). Elaborate research into the performance of SRA was conducted by Vanhoucke (2010a), while Trietsch et al. (2012) investigated the use of lognormal distributions for activity durations. Moreover, several new studies

* Corresponding author at: Faculty of Economics and Business Administration, Ghent University, Tweekerkenstraat 2, 9000 Ghent, Belgium. Tel.: +32 9 264 35 69.
E-mail addresses: jordy.batselier@ugent.be (J. Batselier), mario.vanhoucke@ugent.be (M. Vanhoucke).

http://dx.doi.org/10.1016/j.ijproman.2014.09.004
0263-7863/00/

are concerned with risk analysis in general (Hartono et al., 2014; Nasirzadeh et al., 2014; Taroun, 2014; Zhang and Fan, 2014; Zwikael et al., 2014). Research on project control has also been widespread, with an emphasis on the development of probabilistic forecasting techniques based on the EVM methodology (Caron et al., 2013; Kim and Reinschmidt, 2009, 2010, 2011; Lipke et al., 2009; Naeni et al., 2011; Narbaev and De Marco, 2014). Furthermore, several authors proposed to integrate risk analysis techniques with project control approaches (e.g. SRA with EVM) in order to enhance the decision-making process for taking corrective actions (Acebes et al., 2014; Colin and Vanhoucke, 2014; Elshaer, 2013; Pajares and López-Paredes, 2011; Vanhoucke, 2011, 2012a).

This last research trend has inspired the renaming of the concept of dynamic scheduling to *integrated project management and control* (Vanhoucke, 2014). The new term indeed better represents the integrative interaction between the three dimensions, especially between risk analysis and project control. Furthermore, it stresses the importance of project control. Indeed, controlling the project during its progress should be the main concern of a project manager. Therefore, further discussions on integrated project management and control will mainly refer to the project control dimension, and thus to EVM.

Previous studies have conducted research on integrated project management and control based on both generated (Agrawal et al., 1996; Boctor, 1993; Demeulemeester et al., 2003; Kolisch and Sprecher, 1996; Kolisch et al., 1995; Schwindt, 1995; Tavares, 1999; Van Peteghem and Vanhoucke, 2014; Vanhoucke et al., 2008) and real-life project data (Bright and Howard, 1981; Covach et al., 1981; Hecht, 2007; Henderson, 2003, 2005; Lipke, 2009; Riedel and Chance, 1989; Rujirayanyong, 2009; Tzaveas et al., 2010; Vandevoorde and Vanhoucke, 2006; Zwikael et al., 2000). However, in the latter case, the data set was often of insufficient size and always of inadequate diversity for obtaining generalizable conclusions, as Table 1 indicates. Regarding diversity, the data sets consisted of mutually similar projects situated within the same sector and mostly even within the same company. This is, for example, also the case for the study of Riedel and Chance (1989) listed in Table 1, for which a sufficient number of projects was considered, however, all originating from one and the same US Air Force division.

Note that size and diversity can indeed be identified as the two main quality-determining properties of a database (Christensen et al., 1995; Vanhoucke, 2011). Consequently, many authors have expressed the need for performing integrated project management and control research – in particular research on project control – on a large and diverse data set of real-life projects (Henderson, 2003, 2004, 2005; Lipke, 2009, 2013; Tzaveas et al., 2010; Vanhoucke, 2011; Zwikael et al., 2000). However, no existing project database seems to meet these requirements. Therefore, the main goal and contribution of this paper lies in the creation of a large and diverse database consisting of qualitative real-life project data that can become the basis for multiple studies related to one or more of the dynamic scheduling dimensions. Moreover, to ensure that the project data are indeed qualitative and sufficiently diverse, a framework for database construction and evaluation will be introduced.

Table 1
Chronological overview of project control studies using real-life project data.

Paper	# Projects	# Sectors (companies)
Bright and Howard (1981)	11	1 (1)
Covach et al. (1981)	17	1 (1)
Riedel and Chance (1989)	56	1 (1)
Zwikael et al. (2000)	12	1 (1)
Henderson (2003)	6	1 (1)
Henderson (2005)	1	1 (1)
Vandevoorde and Vanhoucke (2006)	3	1 (1)
Hecht (2007)	1	1 (1)
Lipke (2009)	16 [a]	2 (2)
Rujirayanyong (2009)	2	1 (1)
Tzaveas et al. (2010)	1	1 (1)
This paper (2014)	51	5 [b] (47)

[a] These projects are the same as those used by Zwikael et al. (2000) and Henderson (2003).
[b] There are 5 sectors if construction is perceived as one overarching industry. However, if construction is divided into its constituting subsectors (infra Fig. 1a), 9 different sectors can be identified.

Note that none of the earlier data sets mentioned in Table 1 have been made publicly available. For some of the smallest data sets (Rujirayanyong, 2009; Tzaveas et al., 2010; Vandevoorde and Vanhoucke, 2006), somewhat more extensive project data are provided in the paper itself, but never down to the activity level. In contrast, our entire database – including the specific activity data for all projects – can be consulted at www.or-as.be/research/database (OR-AS, 2014a).

Summarized, the contribution of this paper is fourfold:

- A real-life project database is created, outranking the existing empirical databases from project management literature in both size and diversity. Moreover, the database will continuously be extended and is made publicly available, so that it can become the basis for many future studies related to project management.
- A database construction and evaluation framework is developed, based on the so-called project cards (see Section 2.1). Through implementation of these project cards, the quality of the added data can be guaranteed.
- The project cards can also serve didactical purposes. More concretely, they can be used for introducing students to the three dimensions of integrated project management and control – all incorporated in the project cards (see Sections 2.1.2 to 2.1.4) – in a more practical context.
- The performed database evaluation yields interesting initial results (e.g. on the difference between planned and actual project performance and on EVM forecasting accuracy; see Section 3), which can be further explored in a future research.

The specific outline of the rest of the paper is as follows. In Section 2, the framework for database construction and evaluation based on the project cards is introduced. Furthermore, it is presented how these project cards can be applied in order to ensure qualitative database extension. Section 3 then provides an overview of the general characteristics of the constructed database, based on the information that is included in the project cards. Finally, in Section 4, general conclusions are drawn and actions for further research are suggested.