DEVELOPING A PORTFOLIO DATASET OF REAL PROJECTS FOR PORTFOLIO, PROGRAMME AND PROJECT CONTROL MANAGEMENT RESEARCH

ABSTRACT

This paper introduces a dataset comprising a portfolio of six projects each having a Project Management Baseline (PMB) consisting of a budget derived from a standardised first-principles, bottom-up estimation technique utilising a homogeneous set of resources, both consumable and non-consumable, which are inter-related in a highly complex, multi-dimensional manner with appropriate correlation between quantity, productivity rates and cost rates. The data set also includes detailed time-phased actual costs and progress over the life of each project as well as the time-phased values of revenue claimed for each project. The collection and attribution of 12,139 actual costs and the measurement of progress over a period of just under 109 consecutive weeks is consistent and standardised across all six projects providing a unique differentiator to other datasets.

As the combined project set represents a real-life portfolio of work over the 109 week period, and includes revenue information, the data is valid for research in portfolio, program or project (PPP) management and control, where projects are managed as a group, generally in a standardised manner with a limited supply of shared resources to support a strategic business aim. This is believed to be the first time that data of this nature has been introduced into the PPP management landscape.

The dataset introduced in this research can be used in any future studies on portfolio, programme or project management and studies associated with the control of multiple resource-constrained projects, portfolio contingency allocation and portfolio cash flow analysis.

INTRODUCTION

Having access to actual project data facilitates research on project, program and portfolio (PPP) management processes and techniques to examine hypotheses. Whilst databases with higher level project cost and duration information exist — such as the US Government's Defence Acquisition Executive Summary (DAES) and Selected Acquisition Summary (SAR) — these are not always readily available to researchers. Additionally, the information is often not complete or provided in sufficient detail to undertake research in the first instance or validate research undertaken using artificially generated data sets.

This paper aims to fill this gap and introduces a project portfolio data set of 6 completed projects undertaken between 2011 and 2013, a period of just under 109 weeks, which includes a Project Management Baseline (PMB) for each project, presented in two ways, firstly a client facing manner utilising the client's own work breakdown structure (WBS) to establish budget, client costs and program, and secondly a standardised contractor facing manner that utilises a standardised portfolio WBS (PWBS) that enables project control from a contractor's perspective across the portfolio to be completed in a standardised and repeatable manner. Detailed time-phased actual costs are provided based on daily summaries. Earned values over the life of each project are given at monthly intervals for both the client-facing and contractor-facing manner.

The data set includes a total of 12,139 individual cost records across the six projects during the investigation period, all of which were collected and attributed to work completed in an identical manner. The data also includes a MS Project template that enables an automatic network generation for each project complete with duration and budgets for each task.

The use of the presented data set to researchers is multi-faceted—for example, the base estimates used to determine the PMB for each project can be used independently to generate artificial project data through the probabilistic analysis of time, cost, and resource sensitivities, whilst the actual results of each project can be used to inform probability distributions of possible outcomes to feed back into artificial models for similar project scenarios. The inclusion of project revenue also enables the data set to be useful for research outside PPP management, with detailed cash flow analysis enabling research in business and policy related fields.

The remainder of the paper commences with a literature review identifying the current sources of real-life project data suitable for research purposes. In the absence of published information, the characteristics a good data set should possess for the purposes of PPP research are presented. The data set is then introduced complete with an explanation of the two methods of presenting the data, how the PMB was established, how actual costs were accrued, and progress measured. The paper concludes with the identification of various practical uses of the data, both within PPP research and beyond, the limitations of the current research and suggested research for the future.

LITERATURE REVIEW

Sources of Data for Portfolio, Programme and Project Management Research

Real-life project data sources are rare, with the two most commonly used sources being from the US Government, namely the US DAES report DAU (2013) and the SAR acqnotes (2015). Both data bases were established to summarise the contractual reporting obligations of the US Government's mandatory earned value management requirements.

DAU (2013) has summarised the reporting requirements for US Defence projects since the early 1970s. The reported figures are at the highest project level with some projects and programmes not having a contractual requirement to report past an upper-level percentage of completion and therefore actual final costs may not be recorded in the DAES. The data itself is assumed to be accurate based on the requirement for contractors to be compliant with the US Government mandatory system, and in a blow to open research; the data is not publicly available outside the department without prior approval from the Director, Acquisition and Resource Analyses. Whilst research papers relying on this data are publicly available stating high-level findings, the base data is not, making replication, verification or extension to the research difficult.

The SAR acqnotes (2015) is available for public consumption, however the information in this report is not as detailed and final costs of projects or programmes are not always reported as it is common for The Under Secretary of Defence (Acquisition, Technology, and Logistics) to waive reporting of SAR data at 90% completion of the project or programme or if the project or programme no longer falls within the required category for mandatory reporting. This limits the usefulness of the information for research purposes as the final costs of projects must be extrapolated which introduces another level of uncertainty and complexity. Christensen and Rees (2002) provide further commentary on the difficulty in using the databases for research purposes.

Outside of the two US Government sources, Batselier and Vanhoucke (2015) summarised the 11 known project control studies within the literature that use empirical data sets, observing that none of the base data from these studies has been made publicly available. In recognition of the scarcity of data for research purposes they have introduced a data set OR-AS (2020) which is continuously updated and currently includes 133 non-interrelated projects from five industry sectors with a diverse range of project sizes and durations. Project data within this set is also classified with respect to its completeness and authenticity.

Vanhoucke, Coelho et al. (2016) expand on the use of artificial and empirical project data for research purposes, noting that empirical data is often not structured in a manner suited to the needs of researchers. They attribute this to various circumstances, namely: unavailability of data due to its commercially sensitive nature; data not being collected in a structured manner as the project progresses; data not being updated from a project's starting point; and data being reflective of corrective actions as a project progresses to ensure time and cost objectives are met, however without adequate explanation of the corrective action taken. Notwithstanding this, Vanhoucke, Coelho et al. (2016) summarise that empirical project data has an important role to play in research with the goal being to source *'richer empirical data that can be used in academic research and transformed to controlled artificial data that better reflect reality in order to bring the newly developed methodologies even closer to the needs of professional project management'*

Availability of Project Portfolio Data

A portfolio is defined by PMI (2017) as "a collection of projects, programs, sub-portfolios and operations managed as a group to achieve strategic objectives". The literature (Müller, Martinsuo et al. 2008), (Martinsuo 2013), (Jonas, Kock et al. 2013), (Patanakul 2015), (Kaiser, El Arbi et al. 2015) suggests a multitude of success factors to project portfolio management which can be distilled into the following three main categories: selection of projects for alignment with the organisational strategic intent; decision making tools and processes used within the portfolio to ensure its success and balance; and portfolio reporting.

An extensive review of the literature did not identify any published real-life project portfolio data that would enable the recommendations and hypotheses presented in this research field to be tested.

The Characteristics of a Good Data Set for PPPM Research

An extensive review of the literature did not identify any definitive characteristics of a good data set for research within PPP management; therefore the following eight characteristics are offered as a starting point, based purely on the logic of what is deemed necessary for research and experimentation in the field.

- <u>Standardised baseline determination for all projects.</u> The baselines (budget and durations) for each project should be determined in a standardised manner. This enables replication across the portfolio and changes to be completed globally. One of the main purposes of PPP management research is to test hypotheses which requires sensitivity or 'what if' scenario analysis. In order to complete this, the key aspects of project control – quantum of work, resource costs and availability and productivity must be easily altered for modelling purposes.

- Standardised production or collection techniques. As a project progresses, the data must be produced or collected in a standardised manner which is easily repeatable for future projects. This enables all collected data to be treated in the same manner and allows additional data to be added to the set within a defined framework.
- Data attribution. Because each individual piece of data that is produced or collected will ultimately represent the entire data set, each individual piece must be attributed to a defined, quantifiable and measurable outcome at the lowest possible level, such that outcomes are discrete and definable. Yet, data collection should not be onerous. Attribution gives the data meaning and purpose, so it is possibly the most important characteristic of the data set. For example, the production or collection of cost-related data against a discrete physical quantity of measurable work allows actual unit rates to be established. If cost data is not attributed to discrete quantities or the outcome attributed is at too high a level, the establishment of unit rates becomes problematic which in turn introduces estimation errors and reduces forecast accuracy.
- Produced or collected data must equate to the entire project. For research in a particular area or component of a project, the part collection of data that relates to the area of research is valid for research in that area, such as program or duration research of critical paths, or performance of a particular project element. However, as a portfolio looks holistically at the overall performance of the projects within it, the data collected must be truly representative of the entire project and not just a part or component.
- Portfolio data should be continuous across an extended time period. As a portfolio looks holistically across the multiple constituent projects, the data production or collection process must be continuous across an extended period of time to ensure that there are no gaps. The period of time should extend such that the quantity of records is sufficient for the data set to be considered statistically significant.
- Portfolio data should be produced or collected at consistent time intervals across all projects. The time interval or data date at which data is summarised as projects are planned to progress or actually progress should be consistent. This allows summation at the program or portfolio level to occur with reference to a standard data date. For individual projects, additional data dates can be introduced in between the portfolio data date for the purposes of control, however the portfolio data date should always be honoured.
- Data collected must be easily compared to project baselines. This may appear self-evident; however the method of establishing the initial PMB and the method of tracking and collecting data can be very dis-similar, depending upon the purpose, use or target audience of the data, which is demonstrated in the presented data set within this paper. The method used to establish a PMB should directly relate to the method used to attribute collected data and hence forecast Estimates at Completion for both cost and time.

- Portfolio data should include revenue information. An extensive review of portfolio, program and project management research literature did not reveal any available revenue-related information associated with projects, programs or portfolios. Revenue is an important factor in portfolio research as the structure and spread of costs and margin across each project within the portfolio directly influences the cashflow, and therefore health of the portfolio itself. For example, the ability to generate cash early in a project to cover upfront expenses without physical delivery of work may be the difference between inclusion or exclusion of a project from the portfolio. The ability to accurately model the revenue and cashflow of the existing projects within the portfolio and those proposed projects for inclusion provides a powerful decision-making tool for PPP managers.

THE PORTFOLIO DATA SET

The data set introduced in this paper consists of a portfolio of six projects with PMBs, project progress, cost collection and data attribution completed in accordance with the eight characteristics previously nominated. The data set utilises a homogeneous set of resources, both consumable and non-consumable, across the portfolio which are inter-related in a highly complex, multi-dimensional manner with appropriate correlation between quantity, productivity rates of key resources and cost rates for materials, labour, plant and equipment and sub-contract resources across each project and the entire portfolio.

As projects are generally selected for inclusion into a portfolio based on their ability to meet the strategic aim of an organisation, one of which is normally profitability, the ability to model and time phase cashflow is an important decision-making tool. The data presented here also includes actual costs and revenue for each project which enables future estimates to be informed and the portfolio optimised.

It is these characteristics that set this data set apart from other available data sets consisting of individual unrelated projects without explanation of the PMB creation or data produced from fictitious networks and it is believed to be the first time data of this type has been published within PPP management literature. The full data set is available at www.unsw???.

Selection Criteria for Inclusion of Projects

The projects within the portfolio were completed between 2011 and 2012 with projects ranging between $1M (AUD) and $3.6M (AUD) in value and 4 – 12 months in duration. All work within the projects was completed through self-performance, that is, other than specialised activities the risks associated with resource availability, resource rates and productivity were borne by and managed by the contractor.

All projects in the data set were completed on either a fixed lump sum or fixed schedule of rates basis without incentive schemes for an early finish, but damages payable to the client for late finishes. The criteria used for inclusion into the portfolio data set are the projects must: have commenced no earlier than 2011; be completed; have a minimum tender value of $400,000, have an initial planned duration of at least three months and be completed concurrently as part of the portfolio.

These criteria led to the acceptance of six projects into the data set with the first project commencing in January 2011 and the final project being completed in October 2012. In all the data has been collected over an extended duration of just under 109 weeks, with the six projects representing the following:

1. Tendered value of $9.402M
2. A final revenue value of $11,409M
3. Original combined budget of $9,878M
4. Final total actual cost of $9,402M
5. A total of 12,139 individual cost records;
6. A combined planned duration of XXX project days;
7. A total combined actual duration of 1480 project days.

A summary of all projects within the data set is shown in Table 1. Appendix A contains an example of the full detailed breakdown of an example projects. (Should this be a re-direction to a website?)

Project No.	Project Name	Start	End	Duration (Days)	BAC	AC	Tender Value	Final Value
P01	Regional Airport Car Park	5/05/11	6/11/11	185	$1,965,174	$1,890,995	$1,902,511	$2,280,303
P02	Regional Landfill	10/01/11	12/07/11	183	$1,183,216	$1,088,404	$1,254,712	$1,310,603
P03	Regional Bypass Rd	6/05/11	18/03/12	317	$1,372,473	$1,336,973	$863,858	$1,652,653
P04	Regional Development Road	24/05/11	25/10/11	154	$944,431	$813,335	$1,080,647	$952,647
P05	Rehabilitation – Two Roads	22/08/11	19/05/12	271	$1,160,892	$1,017,113	$1,394,688	$1,355,086
P06	Urban Streets Rehab	11/10/11	15/10/12	370	$3,251,742	$3,138,824	$2,906,024	$3,858,444
				1,480	**$9,877,928**	**$9,285,644**	**$9,402,440**	**$11,409,736**

Table 1 – Data set summary

The PMB for each project was established during the tender phase using first-principles, bottom-up estimation techniques. Projects were then decomposed into standardised task-oriented WBS elements with clearly defined units of measure. Using Quantity Metrics as previously described, the time-phased PV for each cost code was established based on estimated production rates and the project schedule. The summation of all cost code PVs at each planned data date equates to the total PV for that given period, and the overall summation of all PVs equates to the PMB.

As with most real-life cases, the majority of those projects in the data set have had scope changes or variations ordered during the execution of the works. All scope changes and variations have been viewed retrospectively through the increase or decrease in the actual quantities assigned to WBS elements or the introduction of new WBS elements if appropriate to accurately define the final approved scope. This method makes the assumption that the final approved scope of work has always been the approved scope, which is a valid proposition given it is impossible to predict when or if a client will order a scope change or variation.

Traditional Client Based WBS versus a Standardised Portfolio WBS

A WBS is a decomposition of a project into smaller discrete tasks that collectively, fully describe the work to be completed in order to improve reporting and control at the project level. PMI (2017) notes the WBS is "a hierarchical decomposition of the total scope of work to be carried out by the project team to accomplish the project objectives and create the required deliverables". The WBS organises and defines the total scope of the project, and represents the work specified in the current approved project scope statement. There is a great deal of literature available on the topic of the WBS, however readers are directed in the first instance to the Practice Standard for Work Breakdown Structures PMI (2006) which fully describes the WBS, the various typical formats a WBS might take, the characteristics of a high-quality WBS and what a high-quality WBS should be capable of achieving within the project, program or portfolio environment.

Some key characteristics of the WBS nominated by PMI (2006) are that the WBS should: be deliverable oriented; fully describe 100% of the project deliverables or outputs with the deliverables described by the WBS limited in size and definition to allow effective control; be appropriate to allow tracking of costs, resource usage and program performance; and be appropriate for overall control of all activities within the project. PMI (2006) also notes that a WBS does not need to directly align with the project program and the individual elements within the WBS are not intended to be network dependent.

The Client's WBS

It is normal for a client to provide a project WBS to a contractor for them to provide price and program against. The client's WBS is generally oriented in the manner to which the client wishes to control the project and report on progress either internally or externally. WBS structures provided by the client very rarely follow the decomposition of a project that best suits project control and reporting within a contractor's organisation, nor do they generally make direct mention of cost items often referred to as overhead costs, such as a contractor's project management, project supervision, payment of insurances, levies or permits, the hire of site facilities or the provision of staff vehicles and accommodation - just to name a few.

At the portfolio level, as each client WBS will be different, the use of the client's WBS for cost control bespoke resource and cost accrual methods for each project. For this reason, a standardised portfolio WBS (PWBS) has been developed for this data set, based on typical standardised, repeated tasks across all six projects, which enables each project to be described in a manner enabling standardisation across the portfolio simplifying reporting and control.

The PWBS

The PWBS used in the dataset maintains the characteristics nominated by PMI (2006). It fully describes the work at the lowest level without being onerous and it does not rely on or follow project network logic. The underlying principle of the PWBS is twofold, firstly to enable a project team from the outset to easily quantify the amount of consumable resources needed and the tasks required to complete the works. The second is to provide a simple method of tracking progress against the PWBS codes that can be reported daily from site-based personnel. By doing this in a standard manner across the portfolio, individuals can transfer between projects without the need for re-training in administrative functions, and methods for determining percentages complete of works and estimates of cost and time at completion across all projects are standardised.

The PWBS used in the data set is shown at Appendix A and is broken into three main categories namely Materials, Production and Overheads, with each category further broken down to enable each project to be fully described in a standardised manner.

Materials. All materials or consumable resources have their own PWBS code with a unit of measure (UOM) commensurate with how the resource is purchased, and without distinction of which tasks the material is used. This is valid as the cost of purchase and delivery of consumable resources to a project site is independent of the effort required to incorporate them into the work.

Production. Production PWBS codes are effort driven, are based on the use of non-consumable resources and do not include the supply of materials. The UOM associated with the PWBS codes enable direct measurement of productivity from the field without the need for detailed survey or complex measurements.

Overheads. Overhead items are those tasks associated with a project that do not generally form part of the direct costs associated with a specific client WBS. Upon submission by a contractor of a client's WBS, overhead costs are generally spread or apportioned across all WBS items, making tracking and attribution of them using the client WBS a complex task requiring a disproportionate effort from the project staff.

The relationship between the client's WBS and the PWBS, the tasks completed at each stage and the usage of each are summarised in Figure 2.

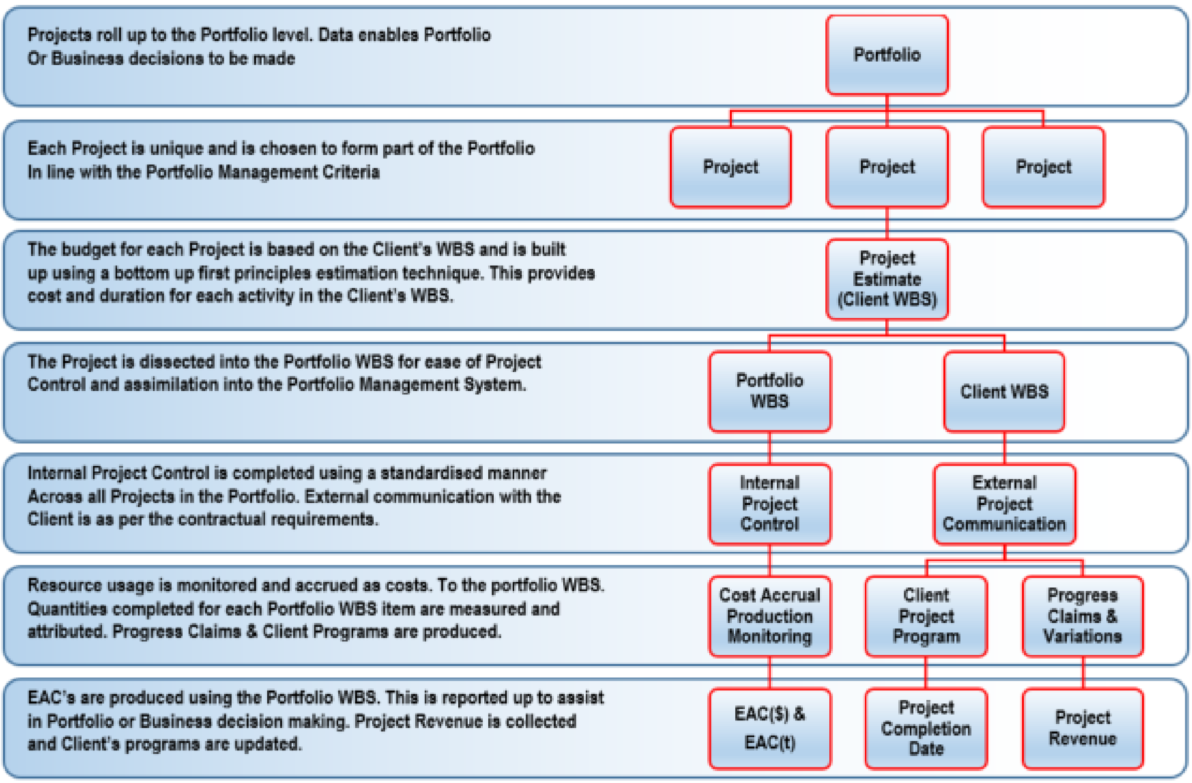


Figure 1 – Relationship and usage of PWBS and client's WBS.

COMPOSITION OF INDIVIDUAL PROJECT DATA

Project Management Baseline of Individual Projects

PMI (2017) notes that Earned Value Management (EVM) techniques for project control are best used in conjunction with a manual bottom-up EAC technique, with the published EAC methods being used at the highest project data level primarily as a comparison tool for calculations against various risk scenarios.

The data is presented in MS Excel Workbook format, chosen due to the ease at which data is exported to other applications or imported as needed. There are six individual files that make up the data set, with each file containing individual worksheets as follows:

Estimate Tab

The estimate tab contains the model for the PMB for each project. The PMB for each project is produced using a bottom up, first principles process. Costs are derived based on the quantity of consumable and non-consumable resources whilst the duration of each task is based on the productivity of non-consumable resources to complete a given quantum of work. The duration of the project is then based on the network logic within the Program Links tab. Time based overheads are then determined from the overall project duration. Within each item. the resources needed to complete the task and their associated productivities are correlated or linked to ensure the validity of model runs if sensitivity analysis is undertaken. For example, if an excavator is loading three trucks in a cycle, the overall time required by the excavator and trucks are linked. If a sensitivity analysis is undertaken, the model will not allow the excavator to speed up, whilst the trucks slow down and vice-versa.

The process of building up the PMB is repeatable across the portfolio, and as the project progresses updates to the PMB are easily made to accommodate scope change and provide actual outcomes at the highest project level for ease of comparison.

The estimate contains a user defined WBS number in column A for items and sub-items within it. This is purely for the purpose of assisting with the look up functions associated with sensitivity analysis. The client's WBS descriptors are contained in column B and C respectively. The estimate contains cells which are coloured green. Each of these cells is a variable which can be changed for the purpose of sensitivity analysis. The costs of resources, both consumable and non-consumable are altered in the resources tab.

Resources Tab

The resources tab contains all consumable and non-consumable resources used to build up the PMB. The resources are separated into four types – labour, plant and sub-contractor which are all non-consumable resources and materials which are consumable resources. The UOM and cost rate for each resource is provided here and can be changed globally. The estimate tab utilises a lookup function to insert the rate of the resource. For sensitivity analysis, rates can be modelled with discrete outcomes directed back to the rate cells to inform individual model runs.

Model Inputs Tab

As previously noted, the estimate contains green cells, each of which is a variable that can be altered for the purpose of sensitivity analysis. The model inputs tab contains all variables, with

UOM noted and the type of the variable, either quantity based or productivity based. Again, for sensitivity analysis quantities and productivity rates can be modelled with discrete outcomes directed back to the model input cells to inform individual model runs.

Non-Workdays Tab

The non-workdays tab contains the non-workday calendar in the locality the projects were completed for the portfolio plus an additional two years from 2011 – 2015 to assist in modelling, the start date for project and the base working hours per day. It is provided to ensure the networks produced are aligned with the timespans of actual delivery and the costs accrued have meaning. The working hours per day is utilised within the estimate to determine the duration of each task and can be changed globally. The portfolio calendar is incorporated into the MS Project template provided with the dataset.

Program Links Tab

The program links tab contains the network logic required to construct the project program, together with the actual progress against each project activity over the life of the project. Copying column A to E into the provided MS Project template and auto-generating the schedule produces a fully linked project network complete with durations and direct cost budgets for each task.

Budget and Revenue Tab

The budget and revenue tab contains the actual direct cost budget and contract rates, totals and agreed progress for the project based on the client's WBS and includes all project variations. For the purpose of generating progress claims, revenue and reporting to the client, the progress of each WBS item was measured and agreed, generally on a monthly basis, which then formed the basis of payment claims and updates to the project program. This tab contains the agreed measurement and revenue paid for the project on a time phased basis.

Portfolio WBS Tab

The Portfolio WBS tab dissects the project into the PWBS format and describes the project in the following manner: The PMB is described via the planned value (PV) over time of the project, with the PV of each PWBS item being the planned quantity (PQ) of the work at the completion of each month multiplied by the PWBS item budget.

The actual progress of the project is shown as the actual quantity (AQ) completed for each PWBS item at the end of each month, with the earned value (EV) being the AQ multiplied by the PWBS item rate. Actual costs (AC) for each PWBS item are summarised on a cumulative basis at the end of each month.

Actual Costs Tab

The actual costs tab is a full account of all costs associated with the project on the date to which they were accrued. They are summarised by PWBS item in date order with the resource, quantity, rates and totals shown.

Actual resource usage and materials received on site were recorded daily on a standardised time sheet and allocated to relevant PWBS items. Once received in the head office, usually within 24 hours, resources and materials were accrued against each PWBS item within the company finance and project control system, with actual costs calculated based on the contract or agreed cost rates for resources and materials.

This standardised method enabled a simple manner of data collection from the field and provided a sound feedback mechanism for the initial first-principles cost-estimation process as project monitoring and control occurred in the same manner used to establish the PMB.

To protect privacy of individuals, names have been removed and replaced with descriptions such as labour, project supervision, project management and project administration. A time phased summary of the AC by PWBS items is also provided.

Post Action

The data set is presented as post action that is as projects progressed, necessary corrective actions were taken to ensure a minimisation of actual costs and maximisation of production and profit. This is a valid position, supported by Vanhoucke, Coelho et al. (2016) who note the main reason data is collected in the first instance is to support the decision making of corrective actions and not for the benefit of researchers.

Data-Scrubbing

In the compilation of this data set each of the 12,139 individual cost entries has been reviewed for accuracy and validity to ensure that they have been allocated against the appropriate PWBS item. Any discrepancies identified have been rectified, such as gravel supply being incorrectly allocated to concrete supply, or a sewer installation sub-contractor being allocated to a water main installation sub-contractor for example. It is acknowledged that because some data entries required re-allocation there is the possibility that others have been missed, potentially leaving 'noise' in the data set and it is therefore not perfect, however it is believed to be as accurate as it can be.

SOME PRACTICAL USES OF THE DATA SET

The intent of releasing this data set to the wider project management community is to assist in the validation of PPP and wider research. The data is presented is a manner that may simplify modelling and simulation, with the actual variables that affect the cost and duration of the projects listed separately. Apart from the obvious uses of confirming research hypotheses in project networks, costs and duration the data set can be used for other practical applications. These may include development of probabilistic risk based contingencies for both planned risks (known risks such as possible changes to quantities, resource cost rates and resource productivity) and un-planned risks (scope changes or unforeseen delays) at both the project and portfolio level; the evaluation of business strategies based on accrued and cash positions and the assessment of investment decisions such as hire, lease or buy of major capital items, or increasing the number of permanent workforce as opposed to a labour hire arrangements.

Other potential areas of research the data set may be useful include, but are not limited to include the following.

Resource Constrained Project Scheduling

Identifying, managing and optimising multiple projects under the pressure of resource constraints is a common project management task, especially within the construction industry. This dataset presents six projects which utilise a common set of consumable and non-consumable resources across each project. Whilst the dataset is presented as six individual projects completed over a 109-week period, the models enable resource constraints to be added at the discretion of the researcher following export of the data into the preferred software package, to test hypotheses and scheduling algorithms.

Contingency – How Much and at What Level Should it be Held - Project or Portfolio?

Whilst it is outside the scope of this paper, a common question within PPP management is at what level should contingency be held and what should the value of this contingency be?

As the individual projects represent a portfolio and are related, risks can be modelled across multiple projects informing the establishment of portfolio contingencies, potentially minimising cash and resources needed in management reserves, whilst the revenue data across the portfolio can be used to optimise cash flow.

Again, the optimisation of portfolio contingency and how it relates to organisational and project-based risk is a topic outside of this paper, but one worthy of further research.

Modelling Likely Variations

Most projects have scope change, and it is common from the outset that contractors will identify possible future scope changes as they work through the budget phase. In some instances, these items are not communicated to the client, as the contractor may use this to provide a lower initial lump sum price, in the knowledge that when the scope changes occur margin will be recovered from this.

To effectively achieve this, a contractor must identify which WBS items to reduce and which to increase the margin, as scope changes are often valued using rates within the client's WBS. Therefore, the work that is likely to increase should have a higher margin and vice versa.

This information enables informed decisions to be made depending upon the risk appetite within the organisation.

Cash Flow

The business adage 'Cash is King' has always been true. From a business perspective it is not practical to have a strong accrued position, yet not be able to pay the bills as they fall due. This data set provides a detailed itemisation of all costs associated with the projects, and the days on which these costs were accrued. For example, cash flow models are easily established from the data which would enable a portfolio or business researcher to evaluate contract terms and conditions for the supply of resources or labour agreements and how this may affect the portfolio or business. For example, a common trend in Australia is for larger companies to settle progress claims 45 days after the end of the month in which they are made. The reality of this is that the portfolio must carry the 30 days of the month in question, plus the 45 days until the bill is paid, so up to 75 days of work without a payment. If the portfolio or business is small and cannot demand equivalent payment terms from their suppliers, a larger management reserve will be needed by the business to carry them through the funding gap. This data set allows modelling of such situations and could be used to assess the suitability of Government policies associated with security of payments to smaller organisations through the application of cash flow analysis.

Earned Value and Earned Schedule Techniques

As noted by PMI (2017), EVM techniques for project control are best used in conjunction with a manual bottom-up EAC technique, with the published PMI (2017) EAC methods being used at the highest project data level primarily as comparison tool for calculations against various risk scenarios. This data set provides actual outcomes at the highest project level which allows various control techniques to be compared.

Generation of Additional Artificial Data

Vanhoucke, Coelho et al. (2016) note one of the real advantages of empirical data is its use in creating statistical distributions to inform artificial data models. As the estimates, or models, used within each project to create the PMB are very detailed and specific, they may be used to generate additional, project sector relevant artificial data sets to aid future research with the accuracy and completeness of the actual data able to inform statistical distributions for likely outcomes associated with resource costs, availability and productivity to assist in the testing of hypotheses.

CONCLUSIONS

The prime objective of this paper is to make the following contributions: firstly to introduce a portfolio data set of real project data, collected over a significant continuous period of time, to augment the limited real project data currently available for the purposes of research, and secondly, to introduce for the first time real project portfolio data into the academic literature which is complete and standardised across all projects within the portfolio.

This paper introduces a set of eight characteristics that are deemed necessary for a good data set for research within PPP management, followed by the portfolio data set itself which contains all eight of these characteristics and was compiled to cover a period of 200 continuous weeks. The paper also introduces the concept of a client's WBS for external communication and reporting by a contractor with the client and a portfolio WBS as a standardised method for internal communication and reporting.

The repeatable method is described for establishing the PMB for each project for both the clients and portfolio WBS, as well as methods of attributing progress and accruing and attributing actual costs. The paper concludes with various practical uses of the data set, both within PPP research and other areas.

Limitations

It is acknowledged that this data set has been derived from the civil construction industry and the standardised PWBS has been developed to best suit the needs of the company in question to simplify project and portfolio control and maximise profitability.

Whilst it is recognised that direct replication of the actual standardised PWBS used in this paper in other scenarios, businesses or industries may not be possible, the concept of decomposing a project into standardised PWBS elements that fully describe the project scope using quantitative techniques such that each WBS element has a unique unit of measure and assigned quantity is achievable in every industry for any project and is limited purely by the imagination and resolve of the project manager.

Suggested Future Research and Planned Publications

As noted previously in this paper, the data set presented opens many possibilities with respect to project, program and portfolio research, all of which are possible future research topics.

This paper publishes the first six projects in the portfolio completed between 2011 and 2012. The entire portfolio comprises 18 projects which meet the selection criteria completed between 2011 and 2015 over a period of 200 weeks. It is intended that the portfolio data will be continuously updated over time until all projects in the portfolio are included. Future research planned by the authors utilising the data set includes - the assessment and validation of estimation techniques for project progress; the assessment of standardised, repeatable bottom-up estimate at completion techniques for cost and duration, in line with recommendations of GAO (2009); and the assessment of validity metrics for estimates of project progress, and estimates at completion for cost and duration.

REFERENCES

acqnotes (2015). "http://acqnotes.com/acqnote/acquisitions/selected-acquisition-report-sar - accessed 20/02/2016."

Batselier, J. and M. Vanhoucke (2015). "Construction and evaluation framework for a real-life project database." International Journal of Project Management **33**(3): 697-710.

Christensen, D. S. and D. A. Rees (2002). "Is the CPI-based EAC a lower bound to the final cost of post A-12 contracts?" The Journal of Cost Analysis & Management **4**(1): 55-65.

DAU (2013). Defense Acquisition Guidebook (Production Date: 28 June 2013).

GAO (2009). GAO Cost estimating and assessment guide: best practices for developing and managing capital program costs, DIANE Publishing.

Jonas, D., A. Kock and H. G. Gemünden (2013). "Predicting Project Portfolio Success by Measuring Management Quality—A Longitudinal Study." IEEE Transactions on Engineering Management **60**(2): 215-226.

Kaiser, M. G., F. El Arbi and F. Ahlemann (2015). "Successful project portfolio management beyond project selection techniques: Understanding the role of structural alignment." International Journal of Project Management **33**(1): 126-139.

Martinsuo, M. (2013). "Project portfolio management in practice and in context." International Journal of Project Management **31**(6): 794-803.

Müller, R., M. Martinsuo and T. Blomquist (2008). "Project portfolio control and portfolio management performance in different contexts." Project Management Journal **39**(3): 28-42.

OR-AS. (2020). "Online consultation of the real-life project database and the

corresponding project cards (continuously updated). URL: http://www.or-as.be/research/database." Retrieved 7/03/2020, 2020.

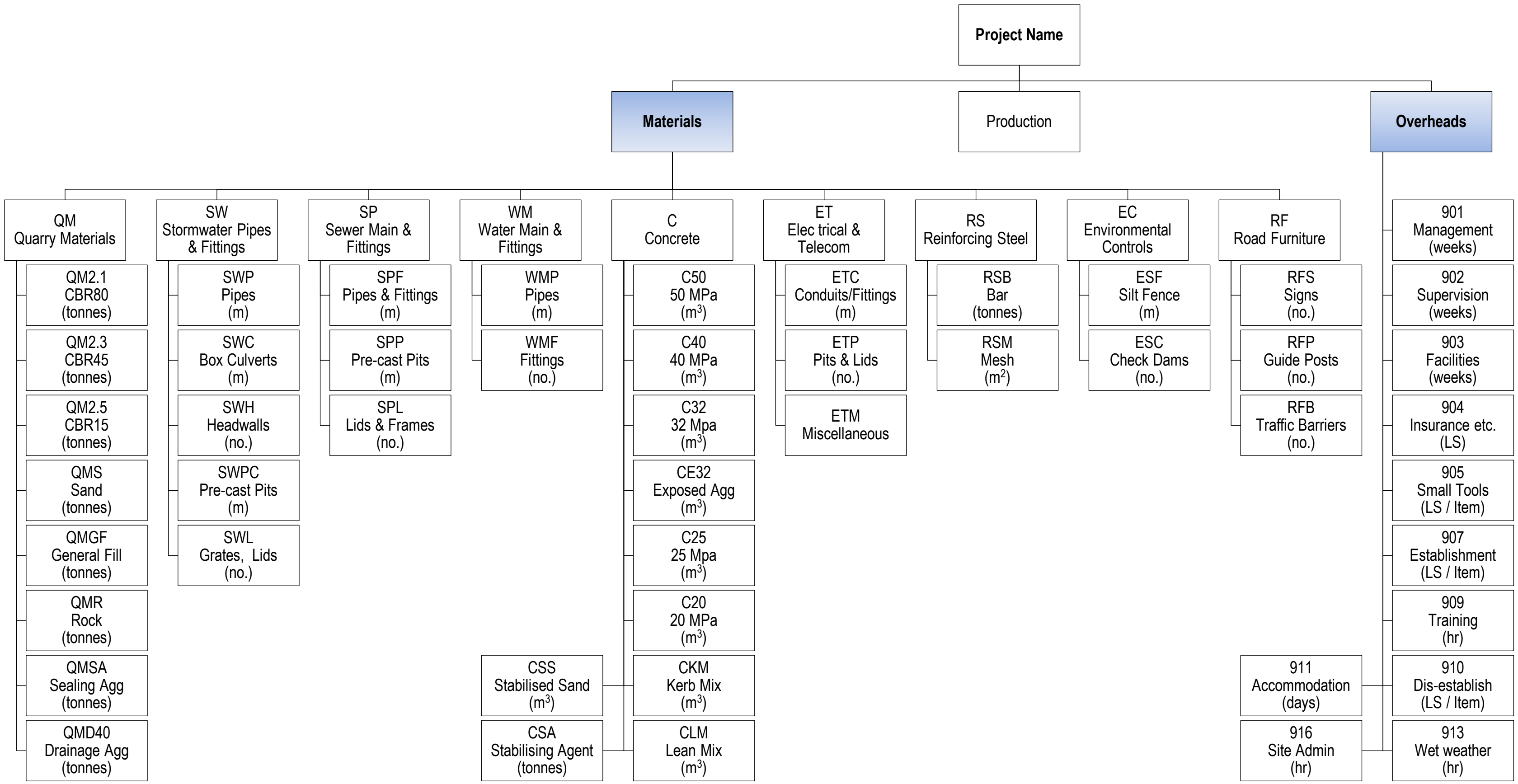
Patanakul, P. (2015). "Key attributes of effectiveness in managing project portfolio." International Journal of Project Management **33**(5): 1084-1097.

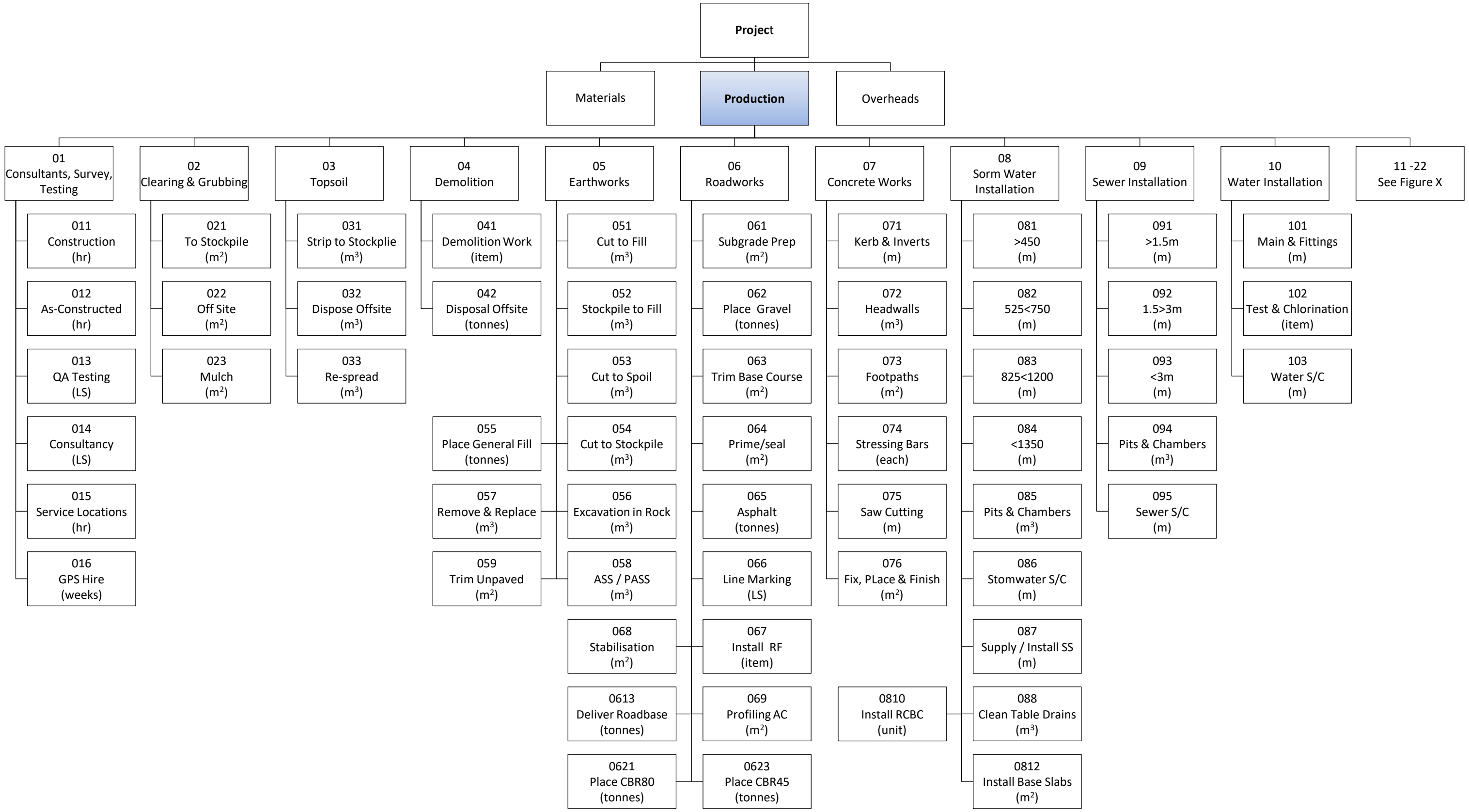
PMI (2006). Practice Standard for Work Breakdown Structures (2nd Edition). Newtown Square, PA, Project Management Institute, Inc. (PMI).

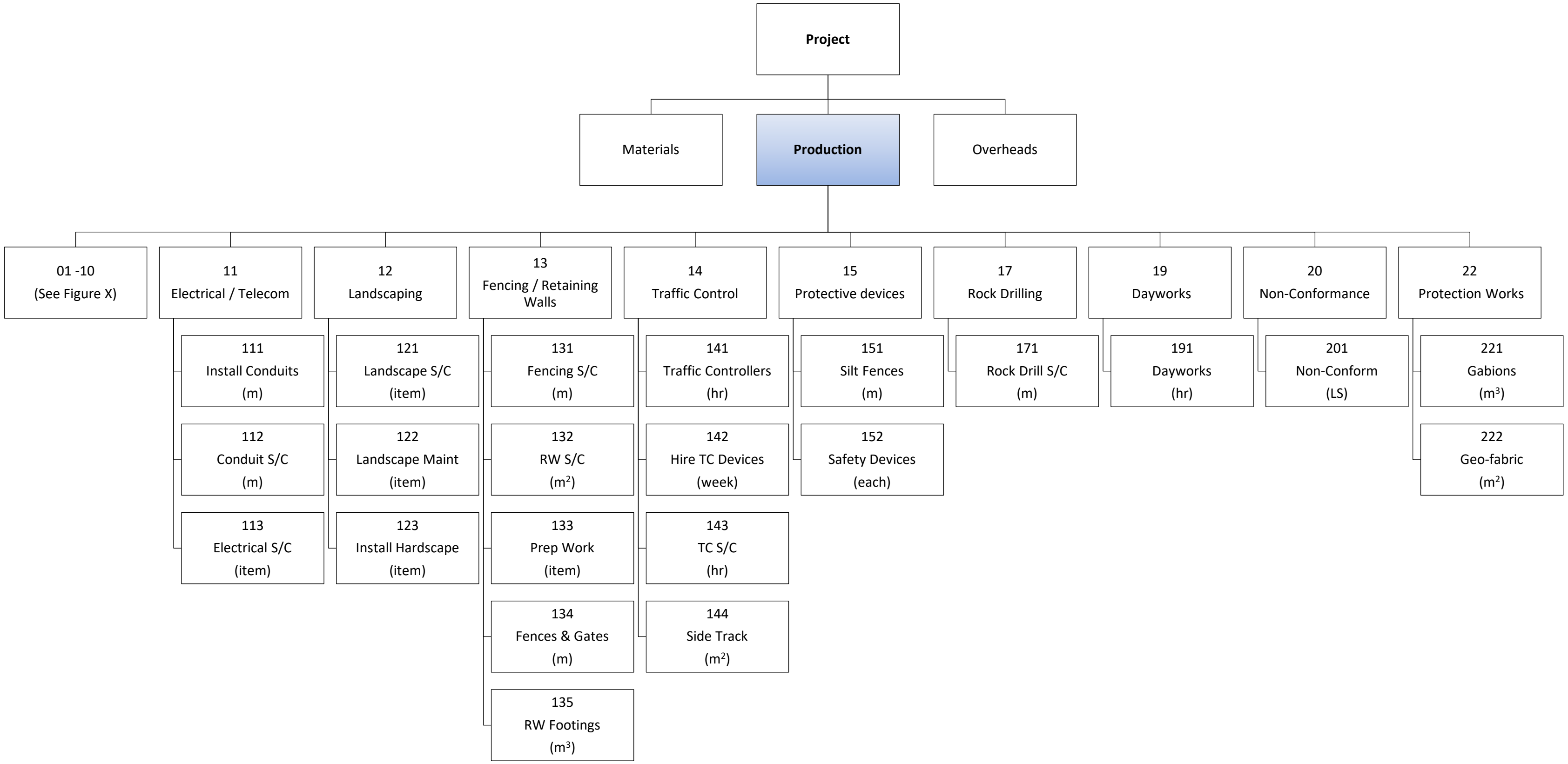
PMI (2017). A guide to the project management body of knowledge (PMBOK guide) - Sixth edition.

Vanhoucke, M., J. Coelho and J. Batselier (2016). "An overview of project data for integrated project management and control." Journal of Modern Project Management **3**(3): 6-21.

APPENDIX A


Project Name
Materials
Production
Overheads
QM Quarry Materials
SW Stormwater Pipes & Fittings
SP Sewer Main & Fittings
WM Water Main & Fittings
C Concrete
ET Elec trical & Telecom
RS Reinforcing Steel
EC Environmental Controls
RF Road Furniture
QM2.1 CBR80 (tonnes)
QM2.3 CBR45 (tonnes)
QM2.5 CBR15 (tonnes)
QMS Sand (tonnes)
QMGF General Fill (tonnes)
QMR Rock (tonnes)
QMSA Sealing Agg (tonnes)
QMD40 Drainage Agg (tonnes)
SWP Pipes (m)
SWC Box Culverts (m)
SWH Headwalls (no.)
SWPC Pre-cast Pits (m)
SWL Grates, Lids (no.)
SPF Pipes & Fittings (m)
SPP Pre-cast Pits (m)
SPL Lids & Frames (no.)
WMP Pipes (m)
WMF Fittings (no.)
CSS Stabilised Sand (m^3)
CSA Stabilising Agent (tonnes)
C50 50 MPa (m^3)
C40 40 MPa (m^3)
C32 32 Mpa (m^3)
CE32 Exposed Agg (m^3)
C25 25 Mpa (m^3)
C20 20 MPa (m^3)
CKM Kerb Mix (m^3)
CLM Lean Mix (m^3)
ETC Conduits/Fittings (m)
ETP Pits & Lids (no.)
ETM Miscellaneous
RSB Bar (tonnes)
RSM Mesh (m^2)
ESF Silt Fence (m)
ESC Check Dams (no.)
RFS Signs (no.)
RFP Guide Posts (no.)
RFB Traffic Barriers (no.)
911 Accommodation (days)
916 Site Admin (hr)
901 Management (weeks)
902 Supervision (weeks)
903 Facilities (weeks)
904 Insurance etc. (LS)
905 Small Tools (LS / Item)
907 Establishment (LS / Item)
909 Training (hr)
910 Dis-establish (LS / Item)
913 Wet weather (hr)


Project
Materials
Production
Overheads
01 Consultants, Survey, Testing
02 Clearing & Grubbing
03 Topsoil
04 Demolition
05 Earthworks
06 Roadworks
07 Concrete Works
08 Sorm Water Installation
09 Sewer Installation
10 Water Installation
11 -22 See Figure X
011 Construction (hr)
012 As-Constructed (hr)
013 QA Testing (LS)
014 Consultancy (LS)
015 Service Locations (hr)
016 GPS Hire (weeks)
021 To Stockpile (m^2)
022 Off Site (m^2)
023 Mulch (m^2)
031 Strip to Stockplie (m^3)
032 Dispose Offsite (m^3)
033 Re-spread (m^3)
041 Demolition Work (item)
042 Disposal Offsite (tonnes)
055 Place General Fill (tonnes)
057 Remove & Replace (m^3)
059 Trim Unpaved (m^2)
051 Cut to Fill (m^3)
052 Stockpile to Fill (m^3)
053 Cut to Spoil (m^3)
054 Cut to Stockpile (m^3)
056 Excavation in Rock (m^3)
058 ASS / PASS (m^3)
068 Stabilisation (m^2)
0613 Deliver Roadbase (tonnes)
0621 Place CBR80 (tonnes)
061 Subgrade Prep (m^2)
062 Place Gravel (tonnes)
063 Trim Base Course (m^2)
064 Prime/seal (m^2)
065 Asphalt (tonnes)
066 Line Marking (LS)
067 Install RF (item)
069 Profiling AC (m^2)
0623 Place CBR45 (tonnes)
071 Kerb & Inverts (m)
072 Headwalls (m^3)
073 Footpaths (m^2)
074 Stressing Bars (each)
075 Saw Cutting (m)
076 Fix, PLace & Finish (m^2)
0810 Install RCBC (unit)
081 >450 (m)
082 525<750 (m)
083 825<1200 (m)
084 <1350 (m)
085 Pits & Chambers (m^3)
086 Stomwater S/C (m)
087 Supply / Install SS (m)
088 Clean Table Drains (m^3)
0812 Install Base Slabs (m^2)
091 >1.5m (m)
092 1.5>3m (m)
093 <3m (m)
094 Pits & Chambers (m^3)
095 Sewer S/C (m)
101 Main & Fittings (m)
102 Test & Chlorination (item)
103 Water S/C (m)


Project
Materials
Production
Overheads
01 -10
(See Figure X)
11
Electrical / Telecom
111
Install Conduits
(m)
112
Conduit S/C
(m)
113
Electrical S/C
(item)
12
Landscaping
121
Landscape S/C
(item)
122
Landscape Maint
(item)
123
Install Hardscape
(item)
13
Fencing / Retaining Walls
131
Fencing S/C
(m)
132
RW S/C
(m²)
133
Prep Work
(item)
134
Fences & Gates
(m)
135
RW Footings
(m³)
14
Traffic Control
141
Traffic Controllers
(hr)
142
Hire TC Devices
(week)
143
TC S/C
(hr)
144
Side Track
(m²)
15
Protective devices
151
Silt Fences
(m)
152
Safety Devices
(each)
17
Rock Drilling
171
Rock Drill S/C
(m)
19
Dayworks
191
Dayworks
(hr)
20
Non-Conformance
201
Non-Conform
(LS)
22
Protection Works
221
Gabions
(m³)
222
Geo-fabric
(m²)

BIOGRAPHY

Brett Thiele is currently a PhD candidate at the University of New South Wales, Canberra and holds bachelor and masters degrees in civil engineering. Brett has over 25 years post graduate experience in project, program and portfolio management, including within the Australian Defence Force and the civil construction industry. Brett is a Fellow of Engineers Australia, an Engineering Executive and Chartered Professional Engineer in the civil college, a Registered Professional Engineer Queensland, a member and Queensland Chapter Councillor of the Australian Institute of Project Management and Certified Practicing Portfolio Executive. Brett is also a member and graduate of the Australian Institute of Company Directors.

Associate Professor Mike Ryan is the Director of the Capability Systems Centre, University of New South Wales, Canberra. He holds bachelor, masters and doctor of philosophy degrees in engineering. In addition, he has completed two years formal engineering management training in the United Kingdom. He has over 35 years of experience in communications engineering, systems engineering, project management, and management. Since joining UNSW, he has lectured in a range of subjects including communications and information systems, systems engineering, requirements engineering and project management and he regularly consults in those fields. He is the author or co-author of 12 books, three book chapters, and over 200 papers. He is a Fellow of Engineers Australia, a Chartered Professional Engineer (CPEng) in electrical and ITEE colleges, a Senior Member of IEEE, a Fellow of the International Council on Systems Engineering, and a Fellow of the Institute of Managers and Leaders.

Dr Alireza Abbasi is a lecturer with the School of Engineering and Information Technology, the University of New South Wales (UNSW Australia), Canberra. He obtained his PhD from the University of Sydney in 2012, MSc at Seoul National University in 2008 and a BSc in Software Engineering at Isfahan University of Technology in 2002. Dr Abbasi has written 1 book, 3 book chapters and over 45 technical journal and conference papers. His research interest covers a wide range of topics in Management Science, Organisation and Network Science and Information Systems Management.