AN ESTIMATION PROCEDURE TO DETERMINE THE EFFORT REQUIRED TO MODEL BUSINESS PROCESSES

Claudia Cappelli[1], Flávia Maria Santoro[1], Vanessa Nunes[1,2]
Márcio de O. Barros[1] and José Roberto Dutra[1]
[1]*Postgraduate Information Systems Program – UNIRIO, Av. Pasteur 458, Urca – Rio de Janeiro, RJ, Brazil*
[2]*COPPE/UFRJ – Systems Engineering and Computer Science Program, Rio de Janeiro, RJ - P.O. Box 68511, Brazil*

Keywords: Business process modeling (BPM), Estimation.

Abstract. Business processes modeling projects are increasingly widespread in organizations. Companies have several processes to be identified and modeled. They usually invest much in hiring expert consultants to do such job. However, they still find no guidelines to help them estimate how much a process modeling project will cost or how long this will take. We propose an approach to estimate the effort required to conduct a BPM project and discuss results obtained from over 50 projects in a large Brazilian company.

1 INTRODUCTION

Business process modeling (BPM) projects are increasingly widespread in organizations, often driven by Business Process Management initiatives (Indulska et al., 2009). The expected benefits from these projects are as follows: (i) Provide tools and analysis contributing to process improvement, focusing on eliminating waste, reducing costs and mitigating risks; (ii) Allow, through process mapping, for learning about "how to perform the work", guaranteeing autonomy to new employees, since knowledge would not be concentrated on older employees; (iii) Facilitate process management, contributing to increase customer satisfaction level.

Well-known project management practices are generally applied in conducting BPM projects. One of the first activities to be undertaken is to identify project scope and establish deadlines. Project duration must be validated according to the effort required to complete modeling and an estimated allocation of professionals over time. Companies have several processes to be identified and modeled. They usually invest much in hiring expert consultants to do such job. However, they still find no guidelines to help them estimate how much a BPM project will cost, how long or how many resources it will take. Differences in process type, goals, performers' profiles, and professionals responsible for modeling render a new project always different from previous ones, thus increasing the uncertainty of such estimates. We argue that, from a historical database of observed efforts invested in BPM projects, the organization can extract indicators closer to reality and, therefore, increase the reliability of its estimates.

Among various process modeling methods, the framework adopted in this work is the workflow-oriented method proposed by Sharp (2001). Several diagrams are used in order to model the business processes, with activities, roles, organizational units, objectives, products, etc.. There are a number of tools and notations available for this purpose. The modeling tool referenced in this work is ARIS (Scheer, 1999) which, among others, includes the following models: Value-added Chain (VAC); Event-driven Process Chain (EPC) and Function Allocation Diagram (FAD).

The aim of this paper is to present an approach for effort estimation on BPM projects and discuss results obtained from real project data from over 50 projects in a large Brazilian company. The paper is structured in the following manner: Section 2 presents business process modeling concepts; Section 3 presents the Estimating BPM Project Effort process and Section 4 argues its limitations; Section 5 presents related work; Section 6 shows conclusions and future work.

Cappelli C., Maria Santoro F., Nunes V., de O. Barros M. and Dutra J. (2010).
AN ESTIMATION PROCEDURE TO DETERMINE THE EFFORT REQUIRED TO MODEL BUSINESS PROCESSES.
In *Proceedings of the 12th International Conference on Enterprise Information Systems - Information Systems Analysis and Specification*, pages 178-184
DOI: 10.5220/0002884201780184

2 BUSINESS PROCESS MODEL DATABASE

Nearly 50 actual BPM projects were developed in a large Brazilian petroleum company for approximately two years, each with its specific goals. These projects were conducted according to an institutionalized method based on the framework presented in section 2 and adherent pattern notation. Teams were generally comprised of a number of part-time modelers and one project manager.

We have recorded effort and cost estimates for all these projects. At the beginning, managers used their experience with projects in other companies and knowledge about the analysts whole would take part in the project to estimate a new project's effort. At a second stage, they began to use data collected from previous projects performed within the company.

2.1 Classifying Business Processes

Thanks to the projects described in the former section, we had a great deal of information on conducting BPM projects. This information included people involved in each project, scope and project scheduling. This data formed a rich database, which we have mined to build the effort estimation model.

However, though the information was available, it was not structured so as to allow immediate analysis: it was distributed along several documents, such as Gantt charts, meeting summaries, and many other types of documentation. Our first step was to collect and organize this data into logical groups and attributes. We have managed to collect information about 48 projects classified into three groups:

- Administrative Projects (ADM): processes related to administrative tasks, performed at offices usually distant from the operational plant. Involve collecting information on market demand, controlling the execution of recurrent inspections and maintenance tasks, organizing training sessions or workshops, collecting and communicating production-related information to high management. Administrative processes usually have a simple workflow, consisting of tasks not described in great detail. Information on 13 administrative processes were collected;
- Operational Projects (TOP): on the flip side of the administrative processes, operational processes are directly related to production and to the daily operation of the production plant. These processes are typically performed by technical personnel, who interact directly with equipment gauges and valves installed in the production site. Operational processes are usually described by huge workflows with very detailed procedures to perform each comprising task, along with exception routes to be followed when the process does not behave as expected. Information on 10 operational processes were collected;
- Technical management projects (TMP): these processes are in the middle ground between the administrative and operational processes. While administrative processes are mostly concerned with clerical activities and reporting, TMP are concerned with production continuity and improvement. They typically involve managing resources required to conduct the operation, tracking new production methods and equipment performance, evaluating new production site performance, and so on. These are distinct from operational processes in the sense that they do not involve directly manipulation of equipment used in production. TMP are usually mid-sized processes, if compared to their peers from the former groups, and are strongly subjected to automation. Information about 25 technical management processes were collected.

2.2 Describing Business Processes

After classifying each BPM project as ADM, TOP or TMP, we collected the following information on each process: Project identifier; Project name; Project class; Business unit; Project start and finish date; Project manager; Analysts; Dedication for each analyst; Project participation start and finish date for each analyst; Number of workflows comprising the process (# EPCs); Number of non-decomposable, atomic activities in the workflows of the process (# FADs); # risks; # indicators; # systems; # business requirements; # business rules; # screens; # equipments; and Interface diagram.

The former attributes were collected for all BPM projects comprising our database. Afterwards, we eliminate outliers due to the following reasons:

- One ADM process had too many activities (# FAD). While the average ADM process has 46 activities, the eliminated one had 183 activities (the second larger ADM process had only 79 activities);
- One TMP process was paused throughout a long time frame. Modeling team changed after this period, and the new team had to learn about the process from the start;
- Two TMP processes were too small (5 and 10

activities) and performed in a very short time frame (about one month each);
- One TOP and two TMP processes were discarded due to strong reuse from other processes.

3 ESTIMATING BPM PROJECT EFFORT

The objective of an effort estimation technique is to determine the number of man-days required to accomplish a certain task. To create a new effort estimation technique, one must rely on the following project management relationship, which describes the dependencies among duration, work to be done, and number of workers.

$$D = W/U \tag{1}$$

where D represents task duration, represented in a time unit; W is the amount of work required for a single worker to accomplish the task, also measured in a time unit; and U represents the number of workers available to participate in the task. Given that D times U is a measure of effort (number of workers for a certain period), we have the equality $E=W$. Based on our project database, we have to build models both for project work and effort. Given completed projects, we fit estimation equations to describe the amount of work for a given project based on its attributes. When considering a new project, we have to estimate its attributes, apply to the work model, and calculate its effort.

In the following sections, we develop a model to describe the work involved in a BPM project (section 4.1), the effort for each project composing our database (section 4.2), and estimation models for our three project classes (sections 4.3 to 4.5).

3.1 Describing Project Work

Since we had many attributes to estimate the amount of work required for a given BPM process, our first initiative was to reduce the volume of data before creating an effort model. To identify which information is more relevant to the estimation procedure, we have interviewed some of projects managers comprising our database. Their feedback was very important, and is summarized below.

Project managers emphasized that the cost of a BPM project would probably be related to the number of activities, the complexity of such activities, and the degree of detail in which these should be described. The number of activities is represented by the number of FAD's in a process. So, this attribute should be part of the effort model. Project managers have also stated that activity complexity was related to process type. This relationship was proved by data, as shown in Table 2, which depicts the average time required to model each activity for each process class (and standard deviation). ADM processes were found to be the hardest to model, while TOP projects were found to be the easiest.

Table 2: Relationship between the time required to perform a BPM project and its number of activities.

Type	$\mu_{Time/FAD}$	$\sigma_{Time/FAD}$
ADM	0,74	0,41
TMP	0,53	0,28
TOP	0,47	0,30

Managers supported this conclusion, mentioning that TOP projects are well-known by a number of people who work directly on the process. Due to the differences, we have decided to separate the effort model according to project type.

Finally, to estimate the degree of detail modeled for each activity in a given project, we have defined a derived attribute from several attributes composing our database. The number of elements (NEL) is a count of distinct complementary information produced by a BPM project, usually related to project goals, which determine what kind of information is to be modeled. This information includes elements described in Section 3.2. For instance, if a project lists the risks and the application systems related to a given process, we count as two elements (NEL=2).

Managers have expressed that the effective count of elements might not be relevant, since it would be directly related to the number of activities. Again, data supported this claim. Table 3 presents the correlation (Spearman's rank order coefficient) between the effective count of elements (ECT) identified by a BPM project and the number of activities in the same process (high correlation for ADM and TOP and moderate correlation for TMP). Since ECT is highly correlated to the number of activities and given the difficulties of having this information *a priori*, we have decided not to take it into consideration in our effort estimation model.

Table 3: Correlation between the effective count of elements (ECT) and the number of activities (FAD).

	ADM	*TMP*	TOP	*Overall*
$\rho_{ECT, FAD}$	85%	65%	87%	78%

Thus, the amount of work to be accomplished in order to execute a BPM project *p* depends on the class of the process under analysis and is based on two variables: FAD_p and NEL_p.

$$W_p = f\,(FAD_p, NEL_p) \qquad (2)$$

Sections 4.3, 4.4, and 4.5 present the analytical formulation of the *f()* function for each process class and the limits of its application. However, to derive these formulations we must compare the amount of work required to accomplish the BPM project to the time that was effectively required to complete the work. The following section addresses estimating the time required for each process.

3.2 Describing Project Effort

To calculate the effort (man-days) invested in each project comprising our database, we have multiplied the number of workers participating in a project by the amount of time during which it was performed.

Our BPM projects were performed by two types of workers: analysts and managers. By analyzing our data, we have perceived an almost constant participation of managers, dedicating about 20% of their work-time for each project. Thus, managing a BPM project can be deemed a constant effort and the varying workforce is described solely by the number of analysts involved in the project.

In our experience, it was common practice to assign an analyst to more than one BPM project at a time. This concurrent work is very important, since BPM projects usually have periods in which the results produced by the analysts are being validated by the client, leaving the team available to work on other processes. On the other hand, we do not have precise data about the dedication of each developer to each project in our database. Instead, we know the periods in which each analyst worked for each process and the fraction of an 8-hour workday that s/he worked during this period for all projects in which the analyst was involved.

To estimate the dedication of each analyst to each BPM project, we have assumed that an analyst assigned to more than one project would equally divide the work time among these projects. Thus, if a part-time analyst (50% of an 8-hour workday) has worked for two projects in a given week, s/he dedicated 25% of a man-week to each project. By doing so, we have calculated a derived attribute for each project in our database: the <u>adjusted number of modelers</u> (AM), calculated as per equation 3 below.

$$AM_p = \frac{1}{f_p - s_p} \cdot \sum_{d=s_p}^{f_p} \cdot \sum_{a=a_1}^{a_N} \frac{dedication_a}{\#\,projects_{a,d}}. \qquad (3)$$

where s_p is the start date for project *p*, f_p is the finish date for project *p*, $dedication_a$ represents the whole dedication (for all projects involved) of a given analyst *a* as a fraction of an 8-hour workday; and $\#projects_{a,d}$ represents the number of projects in which an analyst *a* worked concurrently in day *d*.

Thus, the effort required to accomplish a given BPM project *p*, measured in full-time 8-hour man-days, is calculated by multiplying AM_p by project duration, as presented in equation 4.

$$E_p = (f_p - s_p).AM_p \qquad (4)$$

Given E_p for each project in our database, we can fit proper a *f()* function for each project class, as presented in the following sections.

3.3 Estimating BPM Projects

Prior to estimate a BPM project, there is a need to identify the end-to-end process model, by developing the VAC, in order to provide an overview of the main processes. Then, each macro-process of the VAC is decomposed in individual business processes. To estimate the amount of activities for each individual business process, consider that each process will be detailed in a single workflow. If it is not possible to deduce the amount of activities for each individual process, it can be done by analogy, using the knowledge database, searching for projects with similar characteristics and obtaining an average amount of activities.

3.4 Estimating ADM Projects

Following outlier elimination, we had 12 ADM BPM projects in our database. Table 4 presents summary information on these projects.

Table 4: Summary information on ADM projects.

	AVG	*SD*	*Min*	*Max*
FAD_P	46	20	14	79
NEL_P	4	n/a	2	5
D_P	77	27	42	132
E_P	33	27.6	12.3	104.4
E_P/FAD_P	0.74	0.41	0.25	1.48

The best fit for equation *f()* (see equation 2) for ADM projects was a combination of a third order polynomial over the number of activities and a linear equation of the number of elements (see Equation 5). This equation has shown good fitness for all ADM projects ($R^2 > 91\%$).

$$E_p = 0.0014FAD_p^3 - 0.16FAD_p^2 + 5.8FAD_p - 10.1NEL_p - 3.7 \qquad (5)$$

Thus, our estimation procedure for ADM BPM processes can be summarized as:

- Estimate the number of activities for the project under interest. Usual ADM processes range from 15 to 80 activities;
- Estimate the number of elements to be addressed while modeling process. ADM processes usually describe from 3 to 5 distinct elements (systems, indicators, business requirements/rules, screens, and interface diagrams);
- If the number of activities is lower than 80 and the number of elements is lower than 6, apply equation (5) to estimate the effort required to conduct the project, in man-days. Accept estimations up to $E_p + FAD_p * 0.41$, allowing one standard deviation for project risks;
- If the number of activities is greater than 80 or the number of elements is higher than 5, we are not able to determine a fitting equation. In such cases, the equation (5) may yield inadequate values and an estimation range is acceptable from $0.74 * FAD_p$ to $1.56 * FAD_p$, that is:
  $$\mu_E * FAD_p \leq E_p \leq (\mu_E + 2 * \sigma_E) * FAD_p$$

3.5 Estimating TOP Projects

After outlier elimination, we had 9 TOP projects in our database. Table 5 presents summary information about these projects.

Table 5: Summary information about TOP projects.

	AVG	*SD*	*Min*	*Max*
FAD_P	332	252	37	722
NEL_P	6	n/a	4	7
D_P	146	51	78	229
E_P	129	95	27	301
E_P/FAD_P	0,38	0,09	0,27	0,53

The best fit for equation *f()*for TOP projects was again a combination of a third order polynomial over the number of activities and a linear equation over the number of elements, like equation (5). Such equation has shown good fitness for all TOP projects ($R^2 > 95\%$). However, due to negative parameters in the third and first power of the polynomial, this formulation would yield very low (even negative) estimations for mid-sized processes addressing in few elements. Since this unexpected behavior was promoted due to spikes in process data, we have decided to smooth the observed effort data using a third order averaging process. Table 6 presents summary information about these projects after the smoothing.

Table 6: Summary for TOP projects after smoothing.

	AVG	*SD*	*Min*	*Max*
E_P	105	74	27	217
E_P/FAD_P	0,31	0,03	0,27	0,38

The best fit for equation *f()*after smoothing was a combination of a second order polynomial over the number of activities and a linear equation over the number of elements (equation 6). Such equation has shown good fitness for all TOP projects ($R^2 > 98\%$).

$$E_p = 0.00015 FAD_p^2 + 0.175 FAD_p + 3.04 NEL_b + 3.13 \quad (6)$$

Thus, our estimation procedure for TOP BPM processes can be summarized as follows:

- Estimate the number of activities for the project. Small TOP processes usually have from 50 to 250 activities, whereas large processes have 600 or more;
- Estimate the number of elements to be addressed while modeling. Small TOP processes usually describe from 5 or 6 distinct elements, while large TOP processes address 6 or 7 elements;
- Apply formula (6) to estimate the effort required in man-days. Accept a 10% range to allow room for project risks.

3.6 Estimating TMP Projects

Following outlier elimination, we had 20 TMP projects whose data is presented in Table 7.

Table 7: Summary information on TMP projects.

	AVG	*SD*	*Min*	*Max*
FAD_P	54	25	13	109
NEL_P	3	n/a	2	7
D_P	63	42	16	151
E_P	26	15	8	56
E_P/FAD_P	0.53	0.28	0.16	1.13

The available data for TMP processes was much noisier when compared to data for ADM and TOP processes. Therefore, the best fit for equation *f()*for TMP projects was poor ($R^2 \approx 28\%$). As in the TOP process, we have proceeded smoothing the observed effort data by using a third order averaging process. Table 8 presents summary information about TMP projects after the smoothing.

Table 8: Summary for TMP projects after smoothing.

	AVG	*SD*	*Min*	*Max*
E_P	24	7	10	38
E_P/FAD_P	0,5	0,16	0,25	0,92

The best fit for equation *f()*after smoothing was a power function over the number of activities (7). Equation 7 has shown fair fitness for all TMP projects (R^2>65%).

$$E_p = 2.07FAD_p^{0.62} + 0.72 \quad (7)$$

We have also found that the number of elements has had limited influence in the effort model (correlation between effort estimation error and the number of elements was as small as 4%). Thus, the number of elements is not used in the fitting equation. Our estimation procedure for TMP BPM processes can be summarized as:

- Estimate the number of activities for the project under interest. TMP processes usually have from 20 to 100 activities;
- Apply formula (7) to estimate the effort required to conduct the project, in man-days. Accept a 20% range to allow room for project risks.

4 LIMITATIONS AND LESSONS LEARNED

Estimation models are highly dependent on available data quality. We have spent a long time cleaning the information conveyed in our database and analyzing the best way in which this could be used to derive the models. Nevertheless, the resulting models are still limited by our restricted data: we ar not able to describe a fitting equation for large (> 80 FAD) ADM processes, for ADM processes addressing more than 5 distinct elements; accuracy for small TOP processes is very limited, and data for TMP processes is noisy enough to inhibit a precise model.

The model could be richer if more data were available and if this data were collected more accurately. Thus, data must be collected from new BPM projects and inserted in the database to allow further revisions on equations. Particularly, resource allocation must be collected more precisely, for example, using timesheets.

Finally, our data may be biased because it was collected from projects performed by the same team. Though the analysts have changed considerably throughout the 2 years, managements have remained almost the same. Again, data from new projects can improve equations and clear this potential bias.

5 RELATED WORK

To our knowledge, no published work exists that presents techniques to estimate the effort required accomplishing a BPM project. Some authors have written about the similarities between software and process modeling projects (Vanderfeesten et al., 2008) and compared software and process model metrics (Gruhn and Laue, 2006; Cardoso, 2005).

Louis at al. (1991) talk about people knowledge, Hughes (1996) considers the experts for estimation. Bielak (2000) and Moses and Clifford (2000) discuss the importance of historical data to improvement in estimate accuracy.

6 CONCLUSIONS AND FUTURE PERSPECTIVES

Business process modeling is a key element when discussing organizational management and IT strategies. Nevertheless, estimating the effort to model business processes has become a great challenge due to the lack of guidelines to support such task. We have presented an estimation procedure approach to determine the effort required to model business processes using data collected for approximately two years in about 50 projects. Data about real effort and resources in each project were stored in a database thus rendering possible the use of such information in this work.

By grouping projects in related process types, we have collected information that was registered in an unstructured way. Through a combination of statistical analysis and expert opinions, we have developed estimation models for each process type group, considering the limitations discussed.

As future work, we intend to evaluate the proposed models by applying them on new BPM projects to calibrate the parameters and adjust the formulas, if necessary.

REFERENCES

Bielak J. “Improving size estimates using historical data” In: IEEE Software, Vol.17, Issue 6, p. 27-35, 2000.

Cardoso, J.. How to measure the control-flow complexity of web processes and workflows. In *The Workflow Handbook*, pages 199–212, 2005.

Gruhn, V., Laue, R. Complexity metrics for business process models, in: 9th International Conference on Business Information Systems, Austria, 2006.

Hughes R. T. “Expert judgement as an estimating method” In: Information and Software Technology 38, Elsevier, p.67-75, 1996.

Indulska, M., Recker, J. C. and Rosemann, M., Green, P. *Business process modeling: current issues and future challenges.* International Conference on Advanced

Information Systems, 8-12 2009, The Netherlands.
Louis M. Taff, James W. Borchering, and W. Richard Hudgins, Jr. "Estimeetings: Development Estimates and a Front-End Process For a Large Project" In: IEEE Transactions on Software Engineering, v.17, n8, 1991.
Moses, J. Clifford, J. "Learning how to improve effort estimation in small software development companies" In: Computer Software and Applications Conference (COMPSAC), The 24th Annual International, Taipei, Taiwan, p. 522-527, 2000.
Scheer, A. W., ARIS – Business Process Frameworks, 2nd ed., Springer, 1999.
Sharp, A.; McDermott, P. Workflow Modeling: Tools for Process Improvement and Application Development. Norwood: Artech House, 2001.
Vanderfeesten, I., Cardoso, J., Mendling, J., Reijers, H. A. and van der Aalst, W.M.P., Quality metrics for business process models. In: Fischer, L. (Ed.), BPM and Workflow Handbook, Future Strategies Inc., Lighthouse Point, FL, USA. pp. 179-190, 2008.