CHAPTER 3

RESEARCH METHODOLOGY

3.1 RESEARCH DESIGN

In the area of social science research, different types of research designs exist namely survey descriptive, survey explorative, experimental design, quasi experimental design, chronological or historical design, case study design (case synthesis, case analysis, conjoint analysis) and mathematical models. Each and every design purports to measure the prefixed logical determinants thereby drawing inferences, conclusions, formulating theories and developing models.

Generally, survey descriptive design mainly focuses on describing the features, attributes and benefits without probing, but on the other hand, survey explorative design concentrates more on the latent dimensions rather than the manifestations. Hence, the survey explorative design encompasses different approaches, statistical tools and techniques to measure different variables under study. The experimental design and the quasi experimental design generally intend to measure the relationships between the cause and effect to show how far the variables are causally correlated by using relevant statistical techniques to formulate a theory.

The chronological or historical design is mainly to narrate the events in any one of the following order namely chronological sequences,
individual alignments, context similarities and dissimilarities and based on that, a qualitative inference will be made for the general readers.

The case studies generally intend to synthesize or at times analyse and provide prescriptions related to an individual, an organization, an institution, and towards a larger society.

The mathematical model generally includes scientific methodology namely method of agreement, method of difference, joint method, residual method and concomitant variations. Based on this, a simple to advanced mathematical modelling can be formulated to draw higher order truth.

As far as this study is concerned, the Researcher identified survey explorative design as the most appropriate research design for the study. The survey explorative design normally has the following steps namely identification of the research issue, selection of the research problem, formation of general and specific objectives, formulation of null and alternative hypothesis, identification of the universe, selection of the sample, constructions of research instruments, pilot study, data collection procedure, data analysis (both quantitative and qualitative), and drawing inferences/conclusions.

3.1.1 Identification of the Research Issue

The Researcher has identified “Software Metrics Management Practices” as the major research issue in the contemporary research basically because the entire gamut of operations and the resultant outcome – either success or failure depends on so many factors of which software Metrics Management Practices is an important one. The time and cost aspect which is often the most challenging issue in any software development is dependent largely on the successful implementation of software metrics. Generally, the

3.1.2 Selection of the Research Problem

Once the Researcher identified the issue “Software Metrics Management Practices”, the Researcher analysed through a delimitation process to identify the exact problem to be addressed for the future with contemporary practices followed in the Industry. Both National and International scenario when considered in-depth, followed by several rounds of discussions and consultations with the experts in the Industry, it has come to the limelight that the data available about the In-House Software Development Centres (IHDC) and their performance, practices are very limited. Hence, it was decided to show interest in this area to start the research using an explorative research design to reveal some fundamental observations about the Indian In-House Development Centres and to make the research as an empirical study. However, as it goes with any such research about the population, the In-House Development Centres are also too many in numbers and it is difficult to cover the entire population which is very large and ever growing. Therefore, it was decided to have the research done for a few sectors and hence the following six Sectors namely Education, Engineering, Finance, Health, Logistics, and Tourism have been selected. Based on the preliminary review and informal discussions had with the professionals in the Industry, the Researcher narrowed down the problem to “An Empirical Study on the Software Metrics Management Practices in In-House Development Centres”. After selection of the problem, the formation of objectives was done by the Researcher followed by formulation of null and alternative hypotheses. The
objective and the hypotheses have been presented in Chapter 1 of this research thesis.

3.1.3 Identification of the Universe

The researcher, purely based on expert opinion and judgement, considered to study the universe which is the In-House Software Development Development Centres or otherwise known as In-House Development Centres (IHDC) in India. This is a very large base and very fast growing in nature. As the nature of research is survey explorative, the Researcher chose the delphi way to address the issue and a number of experts in the Industry were consulted and therefore the Researcher identified six major sectors namely Education, Engineering, Finance, Health, Logistics and Tourism for the study although there are a number of other sectors having a large number of IHDCs in the country. The study mainly focuses on the non IT companies in India where software development happens by its own development team for its own use. Many IT companies may develop their own software, but they do not come under the IHDC category. The Researcher defines In-House Development Centre as the one which is a part of a non IT company in India having its own development team to develop software for its own use.

3.1.4 Selection of the Sample and Sampling Technique

As the population size is unknown and the research is survey explorative type which is mostly qualitative in nature, the Researcher after making due consultations with the experts in the field and after going through the relevant literature (Tyson Gingery 2013) on the sampling process chose to administer both the quota sampling to make it heterogeneous and snowball sampling (www.survey.cvent.com 2013) to cover the planned number of Respondents through known contacts who in turn would help to introduce their own contacts to bring out common findings from the sample under
consideration (www.qualitative-research.net 2014). Although determining the sample size requires extensive calculation based on several parameters, as a thumb rule particularly when the research is qualitative or mixed in nature (Mason Mark 2014), the sampling size can be from 20 – 30 (Rob Sheldon 2014). If there are sub groups then for each sub group there need to be 20-30 numbers as the sample size (www.uniteforsight.org 2013). What really matters more is the quality of the samples rather than the quantity alone. However, error estimation at 95% level of confidence could be calculated by the formula $1/\sqrt{N}$, where N stands for the sample size. A sample size of 300 would have an error margin of 5.77% ($1/\sqrt{300}=0.0577$). Based on this, the sample size preferred was a number between 200 and 300. However, the total responses received were 253 out of the 300 planned.

3.1.5 Selection of the Sampling Frame

In this study, the Researcher has followed two types of sampling method namely Quota sampling procedure and Snow ball sampling procedure. The intention behind these two types of sampling is mainly to identify heterogeneous sub groups, here in this case, the six Sectors namely Education, Engineering, Finance, Health, Logistics and Tourism. As the study itself is confidential in nature, the Respondents would open up with the Researcher with their sincere and honest response only when they are contacted with known sources which justifies the reason for choosing the Snow ball sampling procedure. The Researcher identified six Sectors for primary data collection and collected data from 253 In-House Development Heads out of 300 contacted. From Education sector there were 50 responses received, 52 from Engineering, 51 from Finance, 26 from Health, 50 from Logistics and 24 from Tourism sector out of the total 300 contacted.
3.1.6 Construction of the Research Instrument

In any type of research, data is of the type primary and secondary. In this study, the Researcher collected both primary and secondary data for the study purpose. Review of literature was done to collect necessary information about the state of affairs in the selected subject and the same was used for the construction of the research instrument. For collecting primary data, a questionnaire was designed. The questionnaire was tested using a pilot study and thereafter small changes were made in the questionnaire. The Researcher considered all types of questions, particularly both open ended type questions and closed ended type questions. As some of the data were ordinal in nature the three point and four point Likert scales were used to capture the data besides capturing binary responses such as yes or no. The questions framed were in line with the objectives of the study and no unnecessary questions were asked. The Respondents were made to feel at ease by making certain responses as optional due to the nature of the question asked and to ensure secrecy. The questionnaire was constructed in a simple and brief way with simple words as prevailing in the Industry and as understood by the IHDC Heads and it contained four sections, the first section on demographic data, the second about the perception of the IHDC Heads on the organizational performance factors, the third one about the Metrics Management Practices and the fourth one about the organizational factors prevailing and that which leads to success. On the whole the questionnaire was very simple and designed to the point and tested for due deployment.

3.1.7 Pilot Study

The Researcher always believed besides the conceptual understanding, field level feedback is often required to modify the thought process in a rational order. With this in mind, the Researcher tested the instrument in four sectors with 35 in-house development heads and based on
their suggestions and recommendations, the instruments has been suitably modified in certain areas, altered in few aspects and deleted in few portions. This kind of meaningful interaction further strengthened the tool in an advanced fashion. In this context, the Researcher has emphatically pointed out how the original thinking of the Researcher and modified version will bring academic richness.

3.2 DATA COLLECTION AND TOOLS USED

This section describes about the data collection procedure and various data analysis tools used such as Measures of Central Tendency, Chi-Square test, Spearman’s Correlation test, Kruskal Wallis ANOVA test and the like.

3.2.1 Data Collection Procedure

Primary data from 253 In-House Development Heads were collected after sending email questionnaires to around 300 in-house development heads followed by some reminders.

Table 3.1 Data Collection from IHDCs

<table>
<thead>
<tr>
<th>Sectors</th>
<th>Sent</th>
<th>Email/Phone</th>
<th>Direct</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Education</td>
<td>60</td>
<td>42</td>
<td>8</td>
<td>50</td>
</tr>
<tr>
<td>Engineering</td>
<td>60</td>
<td>36</td>
<td>16</td>
<td>52</td>
</tr>
<tr>
<td>Finance</td>
<td>60</td>
<td>41</td>
<td>10</td>
<td>51</td>
</tr>
<tr>
<td>Health</td>
<td>30</td>
<td>21</td>
<td>5</td>
<td>26</td>
</tr>
<tr>
<td>Logistics</td>
<td>60</td>
<td>37</td>
<td>13</td>
<td>50</td>
</tr>
<tr>
<td>Tourism</td>
<td>30</td>
<td>19</td>
<td>5</td>
<td>24</td>
</tr>
<tr>
<td>Total</td>
<td>300</td>
<td>196</td>
<td>57</td>
<td>253</td>
</tr>
</tbody>
</table>

After a few reminders, the Researcher collected data from 196 Heads of In-House Development Centres by email and phone and the
remaining 57 could be completed by directly interacting with them. During the interactions, the Heads requested the Researcher to observe their own department in the researcher’s perspective and this is one of the sources of primary data. Using the primary and secondary data the Researcher could get an in-depth insight into the issues at large with respect to the IHDCs and the same has enabled the Researcher to give recommendations.

3.2.2 Data Analysis Tools

As far as this research study is concerned, the Researcher adopted to survey explorative design by identifying a proper universe, and based on that through appropriate sampling technique a representative sample has been drawn and by using relevant research instrument, intended data has been collected and the same has been analysed with statistical techniques. The collected data from 253 in-house software development heads were coded in two ways namely Pre-coding and Post-coding exercises. The remaining qualitative questions were recorded through verbatim statements and based on that, data collection has been completed (www.marketingdonut.co.uk 2014). All the collected data were analysed through SPSS package and EXCEL. The Researcher used measures of central tendency, chi-square analysis, Spearman’s Rank Correlation and Kruskal Wallis (Single Factor ANOVA) to compute the test statistics for due presentation (Anderson 2011). Necessary care was taken to use the most appropriate tool for the analysis of data and drawing conclusions.

3.2.2.1 Percentage Analysis

According to the author Knapp Thomas (2013), “a percentage is a part of a whole. It can take on values between 0 (none of the whole) and 100 (all of the whole). The whole is called the base”. The author Buchanan (1974) in his article on “Nominal and Ordinal Bivariate Statistics”, discusses
different criteria for a good statistic, and concludes: “The percentage is the most useful statistic ever invented…” Percentage is an alternative method to express a proportion (www.tomswebpage.net. 2013). A percentage is equal to the proportion multiplied by 100. Following formula \((3.1)\) represents the percentage calculation where \(N\) is the total numbers, \(X\) is the selected number for which percentage is being calculated and \(p\) is the percentage value. To denote forty five point two three per cent, it is written as 45.23% (www.statcan.gc.ca 2013)

\[
P = \frac{X \times 100}{N}
\]

(3.1)

In this research, the percentage calculation has been extensively administered on the data for doing percentage analysis followed by interpretation of data.

3.2.2.2 Chi-Square Analysis

The author Cooper (2006) in his book on “Business Research Methods” mentions that a Chi-square test is a non-parametric test of statistical significance for doing a bivariate analysis. According to the author Kowalczyk Devin (2014), it is a non-parametric statistical test used to compare expected data with observed data, and is used to interpret the difference, if any, as statistically significant or not. When it is statistically significant, it may be concluded that the difference is not due to random chance.

The Chi-square test indicates if there is a large difference between collected numbers and expected numbers. If the difference is large, it indicates that there may be something causing a significant change and when a large significant difference is observed, the null hypothesis is rejected meaning that there is no interaction between the variables. Following is the
formula (3.2) used in calculating the Chi-square statistic ($\chi^2$). In the formula ‘$O$’ is the observed value, $E$ is the expected value (www.study.com 2014).

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$ (3.2)

In this research, the Chi-square test was used extensively using SPSS Statistical Tool to test various hypotheses formulated. For all the 2x2 cross tabulation, Yates continuity correction was used for more accurate results and also Fisher’s test statistic wherever appropriate.

### 3.2.2.3 Correlation Analysis

For any non-parametric data, if a correlation test is to be conducted, Spearman Rank Correlation can be administered. In the Spearman’s rank correlation rather than data values, ranks are used. The ranks, of course, satisfy the conditions required for correlation as their relationship is linear, but the Pearson’s Product Moment Correlation test cannot be used on the ranks as the data is not from a bivariate Normal population (psych.unl.edu and www.mei.org.uk 2014). Sometimes the data may be available in the form of ranks. In such a situation as well, the Spearman rank correlation can be used. Following is the formula (3.3) for calculating the Spearman’s rank correlation.

Spearman Rank Correlation Coefficient ($r_S$) = \[ 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \] (3.3)

where $d_i$ is the difference in paired ranks and $n$ is the number of cases. In this research the Researcher has used the Spearman’s rank correlation for doing extensive analysis (Levine 2008).
3.2.2.4 Analysis of Variance – ANOVA

According to the author, the Kruskal-Wallis test is defined as (sometimes also called the “A rank-based nonparametric test that can be used to determine if there are statistically significant differences between two or more groups of an independent variable on a continuous or ordinal dependent variable”). This is also at times called as "one-way ANOVA on ranks". When the data is not normally distributed, the Kruskal-Wallis H test can be applied. It is much less sensitive to outliers and it can be used when a one-way ANOVA is inappropriate (statistics.laerd.com 2014). The Kruskal-Wallis Test was jointly developed by Kruskal and Wallis and hence it is named after them. The Kruskal-Wallis test is distribution free nonparametric test, and is used when the assumptions of ANOVA are not met (www.statisticssolutions.com 2014). The Kruskal-Wallis test statistic can be said to have distribution like a chi-square distribution, with k-1 degrees of freedom where $n_i$ should be greater than 5 (www.statisticssolutions.com 2014). If the calculated value of the Kruskal-Wallis test statistic is more than the critical chi-square value, the null hypothesis can be rejected else the null hypothesis cannot be rejected. ( Heckert Alan 2014).

Following is the formula (3.4) for determining the Kruskal-Wallis test statistic

$$H = \frac{12}{n(n+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(n + 1)$$  \hspace{1cm} (3.4)

where $H$ is the Kruskal Wallis Test Statistic, $n_i (i = 1, 2, \ldots, k)$ represents sample sizes for each of the k groups, $R_i$ is the sum of the ranks of group i.
In this research, the Kruskal Wallis test has been used to find out if there were significant differences among the Sectors in the practice of each of the Metrics Management Categories considered (www.itl.nist.gov 2014).

3.3 SUMMARY OF THE CHAPTER

In this Chapter, the Researcher discussed in detail the Research Methodology, Research Design, Research Issue, Research Problem, Identification of the Universe, Sample Selection and Sampling Technique, Sampling Frame, Construction of the Research Instrument, Details about the Pilot Study, Data Collection Method administered and the Tools used for the Research.