

A Practical Application of Content Analysis
Technical Report No: UL-CSIS-06-3 / Lero-TR-2006-02

Kevin Slowey and Ita Richardson¹
Lero – the Irish Software Engineering Research Centre
and
Dept. of Computer Science and Information Systems
University of Limerick, Ireland

1 Introduction

Content analysis is a technique which can be used to analyse qualitative data. In this report, we present a description of how content analysis was applied to data collected while studying Quality Assurance in Global Software Development.

2 Content Analysis

The content analysis technique used in this project is as described in Miles and Huberman (1994). This technique is designed to analyse a set of field notes to “dissect them meaningfully while keeping the relations between the parts intact”. While the technique is intended for large studies with many different sets of field notes and reports, it can be adapted for use within single case studies such as that presented here.

Miles and Huberman define codes as “tags or labels for assigning units of meaning to the descriptive or inferential information compiled during a study.” The researcher ‘codes’ the data by marking chunks of the text with similar topics or information with the same label, typically some representative word or phrase. It is then possible to cluster, re-label or otherwise sort the information effectively.

The difficult part of this exercise is to decide on the codes with which to label the text. Miles and Huberman specify three classes of code: descriptive, interpretive and pattern codes. Each of these classes places more interpretation on the code than the previous one. As only one case was being dealt with in this instance, the codes used were chiefly descriptive codes with a limited number of interpretative codes.

¹ Kevin Slowey graduated with a M.Sc. in Software Engineering from UL in 2006. He participated in the Siemens Global Studio project sponsored by Lero. Dr. Ita Richardson supervised his dissertation – “A Study of Quality Assurance in Global Software Development”.

3 Documentation available

As a result of the research project which was undertaken, a variety of documentation to be analysed was available.

A primary source of data in this project came from reflective practitioner documentation written by one of the authors who participated on the University of Limerick project team within the Siemens Global Studio Project (GSP) (Richardson et al, 2006). In order to aid this research a log book was written up continuously during the project about the student's involvement. This logbook recorded details of the work, its progress, problems, attitudes within the team and how exterior factors were influencing work on the project. Based on the log book data, a reflective account of the project was written by the participant. This was a useful exercise as it served to order the available material and bring into focus some of the issues arising from the GSP.

The project also generated many documents including meeting minutes and agendas, forum posts and emails. This documentation charted the course of the project in a considerable amount of detail. The project documentation itself included requirements documentation, architecture documentation, coding standards, testing standards and descriptions of acceptance tests.

The total mass of this material was quite significant so content analysis was employed as the analysis technique.

4 Coding the data

As described in Miles and Huberman (1994), a set of codes called a "start list" of codes was generated. In this case, this was done prior to beginning coding although it is recommended that this should be done prior to beginning fieldwork. Because of the nature of this project, this was not practical in this instance². The start list comes from "the conceptual framework, list of research questions, hypothesis, problem areas and/or other key variables that the researcher brings to the study" (Miles and

² When commencing work within the GSP, students were not aware of the aspects of software development that they would be required to undertake. Due to this, dissertation topics were chosen after they commenced the project.

Huberman, 1994). In this case the start list was drawn up from the research questions and the initial impressions generated by reading notes. This list can be seen in Figure 1 below.

Title	Code
QA General	
Process Responsibility Trouble	QAP QAR QAT
QA Attitude	
Me Others	QAAttM QAAttO
QA- how work was done	
Initiate Delegate Respond	QAAI QAAD QAAR
Global	
Advantages Disadvantages	GSDA GSDD
Team	
Interactions- Adrian Interactions- Me Interactions- Central Them and Us Roles	TIA TIM TIC TTaU TR
Motivation	
Central Team Me Others on UL Team	MotCT MotM MotOnT
Organisation	
Research Project Development Project Affected by Student Developers Overhead	ORP ODP OASD OO
Learning	
Process Technology Project Specific	LP LT LPS

Figure 1 Start List of Codes

The text of the notes was examined using these lists and codes were assigned. As expected it was found that some codes were used extensively and others were not

used at all. In accordance with the technique being used the list was refined to allow for this, i.e. popular codes were broken down in subcategories and unused codes were eliminated. The resulting list can be seen in Figure 2. In a larger study this process would have been carried on continuously and refined as the data mounted up from variety of different sources. However in this instance, two revisions was a sufficient degree of complexity as there would be little to be gained from performing further revisions when dealing with what is a relatively small data set. Miles and Huberman (1994) advise writing clear operational definitions for each of the codes. The purpose of this is to ensure that codes are used consistently by different researchers and across different documents being analysed. As in this case there was only one researcher and a relatively small amount of data this was not felt to be necessary.

Category	Title	Code
QA		
	Effective Ineffective Process Points Doing Job Not Doing Job	QAeff QAiff QAPP QADJ QANJ
Team		
	Being Managed Division of Work Them and Us	TBM TDW TTaU
Students		
	Affecting Behaviour Affecting Organisation	SAB SAO
Learning		
	Technology Project Specifics	LTech LProj
QAAttitude		
	Me Team Team (not Me) Central	QAttM QAttT+ QAttT QAttC
Interaction Central Team		
	As Helpful As Boss Disagreement	CTH CTB CTD
Organisation		
	Research Project Development Project	ORP ODP

Figure 2 Final List of Codes

An example of coding is presented in Figure 3. This is a scanned image of a section of text demonstrating how the technique has been applied³. The codes from the first pass, the start list, are on the left hand side of the text. The second pass with the final list of codes is on the right. Analysis was performed on the raw data as this produces more complete results than where a researcher edits notes. If prior editing is carried out, there is a danger that, whether consciously or unconsciously, the final analysis can be influenced.

6 Conclusion

Qualitative data analysis can cause difficulties for researchers if they are not familiar with the technique. This report presents how, in one particular case, data from different sources was analysed to ensure that research findings about Quality Assurance in Global Software Development were presented.

Acknowledgements

The research in this report has been funded by Science Foundation Ireland through the Cluster project GSD for SMEs (Grant no 03/IN3/1408C) within the University of Limerick, Ireland. The MSc in Software Engineering is supported by the Higher Education Authority through the IT Investment Fund.

References

Miles, M.B. and A. M. Hubermann, *Qualitative Data Analysis: An Expanded Sourcebook (2nd Edition)*, 2nd ed: Sage Publications, 1994.

Richardson, Ita, Allen Milweski, Patrick Keil and Neel Mullick, Distributed Development – an Education Perspective on the Global Studio Project, *ICSE 2006 – 28th International Conference on Software Engineering*, Shanghai, China, 20th-28th May 2006, pp 679-684.

³ The text differs slightly from that contained in the original document as it has been proof read and edited with a view to inclusion in this document.