An Evaluation of
Open Source Component Licensing Compliance

Benjamin Peter Alex
MBT (UNSW)

Doctor of Business Administration

December 2007
STATEMENT OF ORIGINALITY

This work contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. I give consent to this copy of my thesis, when deposited in the University Library, being made available for loan and photocopying subject to the provisions of the Copyright Act 1968.

[Signature]
ACKNOWLEDGEMENT OF AUTHORSHIP

I hereby certify that the work embodied in this Thesis is the result of original research, the greater part of which was completed subsequent to admission to candidature for the degree (except in cases where the Committee has granted approval for credit to be granted from previous candidature at another institution).

[Signature]
ACKNOWLEDGEMENTS

There are many people to whom I owe thanks for their support, encouragement and expertise over the past four years of my doctoral program.

I am grateful to Dr Leonard Whitehouse, who acted as my supervisor throughout this research. Through his considerate comments, professional advice and rapid responsiveness, I owe far more than passing thanks to Dr Whitehouse.

Dr Gian Casimir was also instrumental in his support throughout the entire program. From my very first coursework module, right through to assisting during the research phase, Dr Casimir has been ever reliable, helpful and supportive.

Dr Margo Poole, Dr John Burgess and Dr Anton Kriz each provided useful feedback and encouragement on early drafts of the dissertation chapters. Dr Dave Syer was instrumental in helping to test and review the automated collection instrument and online questionnaire. I also thank Dr Rod Johnson and Steven Schuurman, for sparking and maintaining my deep interest in open source. Thanks are also extended to Dr Luke Taylor and Jaime Tarrant for their sustained encouragement, practical support and friendship throughout this journey.

I also wish to thank my family – Mum, Dad and Scott – for their encouragement and support. My mother, Dianne Alex, was especially supportive, proofreading thousands of pages over these past few years.

My final thanks are reserved for my beloved Nina, for the consideration, support, inspiration and love she provided in so many different ways. Tack min älskade.
# TABLE OF CONTENTS

STATEMENT OF ORIGINALITY II  
ACKNOWLEDGEMENT OF AUTHORSHIP III  
ACKNOWLEDGEMENTS IV  
TABLE OF CONTENTS V  
LIST OF TABLES X  
LIST OF FIGURES XIII  
SYNOPSIS XVI  

## 1. CHAPTER 1: INTRODUCTION  
1.1 Background to the Research 1  
1.2 Research Problem and Hypotheses 4  
1.3 Justification for the Research 6  
1.4 Methodology 8  
1.5 Research Plan 11  
1.6 Outline of this Report 12  
1.7 Definitions 13  
1.8 Limitations and Key Assumptions 14  
1.9 Conclusion 15  

## 2. CHAPTER 2: LITERATURE REVIEW  
2.1 Introduction 16  
2.1.1 Methodology 17  
2.2 Parent Theory 19
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.2.1</td>
<td>History</td>
<td>19</td>
</tr>
<tr>
<td>2.2.2</td>
<td>Community</td>
<td>23</td>
</tr>
<tr>
<td>2.2.3</td>
<td>Engineering</td>
<td>25</td>
</tr>
<tr>
<td>2.2.4</td>
<td>Economics</td>
<td>29</td>
</tr>
<tr>
<td>2.3</td>
<td>Research Problem Theory</td>
<td>33</td>
</tr>
<tr>
<td>2.3.1</td>
<td>Importance of Licensing</td>
<td>33</td>
</tr>
<tr>
<td>2.3.2</td>
<td>Copyright and Copyleft</td>
<td>34</td>
</tr>
<tr>
<td>2.3.3</td>
<td>Legal Enforceability</td>
<td>36</td>
</tr>
<tr>
<td>2.3.4</td>
<td>Licensing in Practice</td>
<td>39</td>
</tr>
<tr>
<td>2.4</td>
<td>Theoretical Framework</td>
<td>40</td>
</tr>
<tr>
<td>2.4.1</td>
<td>Theoretical Framework Development Methodology</td>
<td>40</td>
</tr>
<tr>
<td>2.4.2</td>
<td>Proposed Framework</td>
<td>41</td>
</tr>
<tr>
<td>2.4.3</td>
<td>Research Questions</td>
<td>44</td>
</tr>
<tr>
<td>2.4.4</td>
<td>Scope Limitations</td>
<td>47</td>
</tr>
<tr>
<td>2.5</td>
<td>Conclusion</td>
<td>49</td>
</tr>
<tr>
<td>3.1</td>
<td>Introduction</td>
<td>51</td>
</tr>
<tr>
<td>3.2</td>
<td>Research Philosophy and Paradigms</td>
<td>52</td>
</tr>
<tr>
<td>3.2.1</td>
<td>Philosophical Paradigms</td>
<td>52</td>
</tr>
<tr>
<td>3.2.2</td>
<td>Present Research</td>
<td>54</td>
</tr>
<tr>
<td>3.3</td>
<td>Research Design and Method</td>
<td>55</td>
</tr>
<tr>
<td>3.3.1</td>
<td>Theoretical Framework and Variables</td>
<td>55</td>
</tr>
<tr>
<td>3.3.2</td>
<td>Instruments</td>
<td>56</td>
</tr>
<tr>
<td>3.3.3</td>
<td>Development Phases</td>
<td>58</td>
</tr>
<tr>
<td>3.3.4</td>
<td>Exploratory Phase</td>
<td>59</td>
</tr>
<tr>
<td>3.3.5</td>
<td>Pilot Phase</td>
<td>59</td>
</tr>
<tr>
<td>3.3.6</td>
<td>Data Collection Phase</td>
<td>61</td>
</tr>
<tr>
<td>3.3.7</td>
<td>Analysis Phase</td>
<td>61</td>
</tr>
<tr>
<td>3.4</td>
<td>Data Gathering Techniques</td>
<td>61</td>
</tr>
<tr>
<td>3.4.1</td>
<td>Automated Collection Instrument</td>
<td>61</td>
</tr>
<tr>
<td>3.4.2</td>
<td>Online Questionnaire</td>
<td>63</td>
</tr>
<tr>
<td>3.4.3</td>
<td>Simulator</td>
<td>66</td>
</tr>
<tr>
<td>3.5</td>
<td>Administration Methods</td>
<td>66</td>
</tr>
</tbody>
</table>
### 3.6 Sampling
- 3.6.1 General 68
- 3.6.2 Pilot Phase 69
- 3.6.3 Data Collection Phase 69

### 3.7 Limitations and Key Assumptions 71

### 3.8 Ethical Considerations 73

### 3.9 Conclusion 74

### 4. CHAPTER 4: RESULTS 76

#### 4.1 Introduction 76

#### 4.2 Simulator 77
- 4.2.1 Justification 77
- 4.2.2 Design and Input Data 78
- 4.2.3 Testing and Operation 80
- 4.2.4 Export and Data Transformation 81

#### 4.3 SPSS File Structure 83

#### 4.4 Descriptive Statistics 84
- 4.4.1 Nominal Variables 85
- 4.4.2 Ordinal Variables 115
- 4.4.3 Interval Variables 122
- 4.4.4 Ratio Variables 133

#### 4.5 Inferential Statistics 144
- 4.5.1 Hypothesis 1 145
- 4.5.2 Hypothesis 2 149
- 4.5.3 Hypothesis 3 151
- 4.5.4 Hypothesis 4 152
- 4.5.5 Hypothesis 5 154
- 4.5.6 Hypothesis 6 156
- 4.5.7 Hypothesis 7 157
- 4.5.8 Hypothesis 8 160
- 4.5.9 Hypothesis 9 161
- 4.5.10 Hypothesis 10 163

#### 4.6 Conclusion 165
5. **CHAPTER 5: CONCLUSION**

5.1 **Introduction**

5.2 **Research Question Conclusions**

5.2.1 Research Question 1

5.2.2 Research Question 2

5.2.3 Research Question 3

5.2.4 Research Question 4

5.2.5 Research Question 5

5.3 **Research Problem Conclusions**

5.4 **Implications for Theory**

5.5 **Implications for Policy and Practice**

5.6 **Limitations**

5.7 **Implications for Further Research**

6. **GLOSSARY**

REFERENCES

WORD COUNT

ANNEXURE A: PARTICIPANT INVITATION EMAIL

ANNEXURE B: PARTICIPANT INVITATION LETTER

ANNEXURE C: PARTICIPANT INSTRUCTIONS

ANNEXURE D: ONLINE QUESTIONNAIRE FORMAT

ANNEXURE E: ETHICS APPROVAL LETTER

ANNEXURE F: DATA COLLECTION INSTRUMENT

- com.acegitech.dba.introspector.CsvWriter
- com.acegitech.dba.introspector.DependencyDetectionVisitor
<table>
<thead>
<tr>
<th>Class</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>com.acegitech.dba.introspector.DependencyIntrospector</td>
<td>270</td>
</tr>
<tr>
<td>com.acegitech.dba.introspector.DetectedJarInformation</td>
<td>273</td>
</tr>
<tr>
<td>com.acegitech.dba.introspector.DetectedTypeUsage</td>
<td>275</td>
</tr>
<tr>
<td>com.acegitech.dba.introspector.Introspector</td>
<td>277</td>
</tr>
<tr>
<td>com.acegitech.dba.introspector.JarIntrospector</td>
<td>280</td>
</tr>
<tr>
<td>com.acegitech.dba.introspector.LicenseDetector</td>
<td>285</td>
</tr>
<tr>
<td>com.acegitech.dba.introspector.runner.JunitRunner</td>
<td>288</td>
</tr>
<tr>
<td>com.acegitech.dba.introspector.runner.StandaloneRunner</td>
<td>288</td>
</tr>
<tr>
<td>com.acegitech.dba.introspector.util.CollectionUtils</td>
<td>288</td>
</tr>
<tr>
<td>com.acegitech.dba.introspector.util.DigestUtils</td>
<td>293</td>
</tr>
<tr>
<td>com.acegitech.dba.introspector.util.Hex</td>
<td>296</td>
</tr>
<tr>
<td>com.acegitech.dba.introspector.util.ObjectUtils</td>
<td>300</td>
</tr>
<tr>
<td>com.acegitech.dba.introspector.util.StreamUtils</td>
<td>314</td>
</tr>
<tr>
<td>com.acegitech.dba.introspector.util.StringUtils</td>
<td>315</td>
</tr>
</tbody>
</table>

**ANNEXURE G: ENTITY RELATIONSHIP DIAGRAM** 334

**ANNEXURE H: METHOD INVOCATIONS BY COMPONENT** 335

**ANNEXURE I: FLOWCHART SYMBOL REFERENCE** 347
LIST OF TABLES

Table 1.5.1: Research Phases and Key Dates ........................................................ 12
Table 2.4.1: Open Source License Frequency ....................................................... 48
Table 3.2.1: Summary of Mixed Methods Types .................................................. 54
Table 4.2.1: License Abbreviations Adopted in Results ...................................... 82
Table 4.2.2: License Restrictiveness Categories .................................................. 83
Table 4.3.1: SPSS File Structure ........................................................................ 83
Table 4.4.1: Project License ................................................................................ 86
Table 4.4.2: Project Country ............................................................................. 87
Table 4.4.3: Project Externally Deployed ............................................................. 89
Table 4.4.4: Project Distributed in Source Form ................................................. 90
Table 4.4.5: Project Distributed in Binary Form .................................................. 91
Table 4.4.6: Project Distributed in Source or Binary Form ................................. 92
Table 4.4.7: Component License ....................................................................... 93
Table 4.4.8: Compliance Approach ................................................................... 94
Table 4.4.9: Overall Compliance ....................................................................... 95
Table 4.4.10: Modification of Component .......................................................... 96
Table 4.4.11: User Interface Acknowledgement of Component ......................... 97
Table 4.4.12: Documentation Acknowledgement of Component ....................... 98
Table 4.4.13: Coupled Usage of Component Expressly Allowed ....................... 99
Table 4.4.14: Commercially Licensed Open Source Component ....................... 100
Table 4.4.15: Redistribution of Component Source Code .................................. 101
Table 4.4.16: Redistribution of Component Binary Files .................................. 102
Table 4.4.17: Written Offer to Provide Component Source Code ..................... 103
Table 4.4.18: Changes to Component Recorded ............................................... 104
Table 4.4.19: Disclaimer in Project License as Required by Component .......... 105
Table 4.4.20: Country Prohibited from Using Component ............................... 106
Table 4.4.21: Patent Litigation Activity ............................................................... 107
Table 4.4.22: Assessment of Component Redistribution Requirement .............. 108
Table 4.4.23: Assessment of Acknowledgement Requirement ......................... 109
Table 4.4.24: Assessment of Changes Recorded Requirement ......................... 110
Table 4.4.25: Assessment of Project License Disclaimer Requirement .......... 111
Table 4.4.26: Assessment of Country Permitted to Use Requirement ......... 112
Table 4.4.27: Assessment of Patent Proceeding Requirement ..................... 113
Table 4.4.28: Assessment of Compatible License for Redistribution ........... 114
Table 4.4.29: Project License Category Frequencies ................................ 116
Table 4.4.30: Project License Category Statistics ........................................ 118
Table 4.4.31: Project License Category Tests of Normality ....................... 118
Table 4.4.32: Transformed Project License Category Statistics .................. 118
Table 4.4.33: Component License Category Frequencies ................................ 119
Table 4.4.34: Component License Category Statistics ........................................ 121
Table 4.4.35: Component License Category Tests of Normality .................... 121
Table 4.4.36: Awareness of Licensing Issues Frequencies ............................ 123
Table 4.4.37: Awareness of Licensing Issues Statistics ................................. 125
Table 4.4.38: Awareness of Licensing Issues Tests of Normality .................. 125
Table 4.4.39: Confidence of Licensing Compliance Frequencies ..................... 126
Table 4.4.40: Confidence of Licensing Compliance Statistics ......................... 128
Table 4.4.41: Confidence of Licensing Compliance Tests of Normality ........... 128
Table 4.4.42: Derivative Work Score Frequencies ........................................ 129
Table 4.4.43: Derivative Work Score Statistics ............................................. 131
Table 4.4.44: Derivative Work Score Tests of Normality ................................. 131
Table 4.4.45: Transformed Derivative Work Score Frequencies ...................... 131
Table 4.4.46: Reflective References by Component Frequencies .......................... 134
Table 4.4.47: Reflective References by Component Statistics .......................... 137
Table 4.4.48: Reflective References by Component Tests of Normality ............ 137
Table 4.4.49: Method Invocations by Component Statistics .......................... 140
Table 4.4.50: Method Invocations by Component Tests of Normality ............... 140
Table 4.4.51: Project Compliance Ratio Frequencies ...................................... 142
Table 4.4.52: Project Compliance Ratio Statistics .......................................... 144
Table 4.5.1: Weiss (2005) Study License Category Frequencies .................... 146
Table 4.5.2: Project License Category * Study Chi-Square Test Results ........... 148
Table 4.5.3: Project License Category * Study Crosstabulation .................... 149
Table 4.5.4: Licensing Awareness * Confidence Symmetric Measures ............. 151
Table 4.5.5: Awareness of Licensing Issues Homogeneity of Variances .......... 151
Table 4.5.6: Awareness of Licensing Issues Ranks ........................................ 152
Table 4.5.7: Awareness of Licensing Issues Test Statistics .......................... 152
Table 4.5.8: Awareness of Licensing Issues Group Statistics ....................... 152
Table 4.5.9: Confidence of Licensing Compliance Homogeneity of Variances 153
Table 4.5.10: Confidence of Licensing Compliance Ranks ......................... 153
Table 4.5.11: Confidence of Licensing Compliance Test Statistics ............... 154
Table 4.5.12: Confidence of Licensing Compliance Group Statistics .......... 154
Table 4.5.13: Compliance * Project License Category Crosstabulation .......... 155
Table 4.5.14: Compliance * Project License Category Symmetric Measures ... 156
Table 4.5.15: Component Category * Project Category Crosstabulation ........ 157
Table 4.5.16: Component Category * Project Category Symmetric Measures ... 157
Table 4.5.17: Open Source Components Per Project Frequencies ............... 158
Table 4.5.18: OSS Components * Project Category Crosstabulation ............. 159
Table 4.5.19: OSS Components * Project Category Symmetric Measures ....... 160
Table 4.5.20: Extent of Derivation * Project Category Crosstabulation ........ 161
Table 4.5.21: Extent of Derivation * Project Category Symmetric Measures .... 161
Table 4.5.22: Project Method Invocation * Derivation Symmetric Measures .... 162
Table 4.5.23: Reflection * Project License Category Crosstabulation .......... 164
Table 4.5.24: Reflection * Project License Category Symmetric Measures ...... 164
Table 5.1.1: Significance Levels ................................................................ 169
Table 5.3.1: Hypotheses Results .................................................................. 180
Table 5.7.1: Method Invocations by Component Frequencies ..................... 335
LIST OF FIGURES

Figure 1.2.1: OSS Component License Evaluation Framework .......................5
Figure 1.4.1: Key Variables ...........................................................................9
Figure 1.4.2: Key Instruments and their Relationships .................................10
Figure 2.1.1: Literature Map .........................................................................18
Figure 2.4.1: Original OSS Component Licensing Evaluation Framework ......42
Figure 3.3.1: Key Variables ...........................................................................56
Figure 3.3.2: Key Instruments and their Relationships .................................57
Figure 3.4.1: Online Questionnaire Process ..................................................64
Figure 3.5.1: Internet Questionnaire Issues ....................................................67
Figure 3.6.1: Proportionate Sampling Process ...............................................70
Figure 4.2.1: Full Component Information Source ........................................80
Figure 4.4.1: Project License .........................................................................86
Figure 4.4.2: Project Country .........................................................................88
Figure 4.4.3: Project Externally Deployed ......................................................89
Figure 4.4.4: Project Distributed in Source Form ...........................................90
Figure 4.4.5: Project Distributed in Binary Form ..........................................91
Figure 4.4.6: Project Distributed in Source or Binary Form .........................92
Figure 4.4.7: Component License .................................................................93
Figure 4.4.8: Compliance Approach ..............................................................94
Figure 4.4.9: Overall Compliance .................................................................95
Figure 4.4.10: Modification of Component ....................................................96
Figure 4.4.11: User Interface Acknowledgement of Component ....................97
Figure 4.4.12: Documentation Acknowledgement of Component ..................98
Figure 4.4.13: Coupled Usage of Component Expressly Allowed .................99
Figure 4.4.14: Commercially Licensed Open Source Component ...............100
Figure 4.4.15: Redistribution of Component Source Code ..........................101
Figure 4.4.16: Redistribution of Component Binary Files .............................102
Figure 4.4.17: Written Offer to Provide Component Source Code ...............103
Figure 4.4.18: Changes to Component Recorded ..........................................104
Figure 4.4.19: Disclaimer in Project License as Required by Component .......105
Figure 4.4.20: Country Prohibited from Using Component..............................106
Figure 4.4.21: Patent Litigation Activity..........................................................107
Figure 4.4.22: Assessment of Component Redistribution Requirement ............108
Figure 4.4.23: Assessment of Acknowledgement Requirement........................109
Figure 4.4.24: Assessment of Changes Recorded Requirement........................110
Figure 4.4.25: Assessment of Project License Disclaimer Requirement ............111
Figure 4.4.26: Assessment of Country Permitted to Use Requirement...............112
Figure 4.4.27: Assessment of Patent Proceeding Requirement........................113
Figure 4.4.28: Assessment of Compatible License for Redistribution...............114
Figure 4.4.29: Project License Category Histogram .........................................116
Figure 4.4.30: Project License Category Normal Q-Q Plot ................................117
Figure 4.4.31: Component License Category Histogram .....................................119
Figure 4.4.32: Component License Category Normal Q-Q Plot .........................120
Figure 4.4.33: Awareness of Licensing Issues Histogram ....................................123
Figure 4.4.34: Awareness of Licensing Issues Normal Q-Q Plot........................124
Figure 4.4.35: Confidence of Licensing Compliance Histogram .........................126
Figure 4.4.36: Confidence of Licensing Compliance Normal Q-Q Plot ...............127
Figure 4.4.37: Derivative Work Score Histogram ................................................129
Figure 4.4.38: Derivative Work Score Normal Q-Q Plot ....................................130
Figure 4.4.39: Transformed Derivative Work Score Histogram .........................132
Figure 4.4.40: Transformed Derivative Work Score Normal Q-Q Plot ...............132
Figure 4.4.41: Reflective References by Component Histogram .........................135
Figure 4.4.42: Reflective References by Component Normal Q-Q Plot ...............136
Figure 4.4.43: Method Invocations by Component Histogram ...........................139
Figure 4.4.44: Method Invocations by Component Normal Q-Q Plot ...............139
Figure 4.4.45: Method Invocations by Component (log) Histogram ...................141
Figure 4.4.46: Method Invocations by Component (log) Normal Q-Q Plot ..........141
Figure 4.4.47: Project Compliance Ratio Histogram ...........................................143
Figure 4.4.48: Project Compliance Ratio Normal Q-Q Plot ................................144
Figure 4.5.1: Project License Categories from Weiss Study Histogram ...............147
Figure 4.5.2: Project License Categories from Present Study Histogram .............147
Figure 4.5.3: Licensing Awareness * Confidence Complying Scatterplot............150
SYNOPSIS

The increasing modularisation of software has witnessed open source software increasingly deployed as components within software projects. Whether such projects are themselves open source or commercial, the reliance upon open source components brings with it specific licensing requirements. This research evaluates how projects comply with those component licensing requirements.

Based on a quantitative analysis of 619 open source component usages across 111 proprietary and open source software projects, key relationships between licensing compliance, user awareness, project license restrictiveness, component choice and engineering coupling were identified and statistically examined. A theoretical framework for evaluating open source component licensing compliance was developed, along with three custom research instruments and an ordinal license restrictiveness categorization scale. Bytecode instrumentation of Java applications was used to quantitatively gather dependency information and coupling metrics.

Strong statistical evidence demonstrated that projects tend to use components of the same license category, with restrictively licensed projects using greater numbers of open source components. Restrictively licensed projects also more frequently created derivative works, with non-restrictively licensed projects commonly employing reflective techniques to mitigate reciprocal obligations.

Overall, 75% of component usages complied with the license, with less restrictively licensed projects achieving greater compliance. Strong awareness of licensing-related issues was found among users, with most having an awareness of licensing issues and correctly self-assessing their compliance with licenses. Specific trends in non-compliance were identified, with an absence of required acknowledgement and project license disclaimers representing common issues.

There was a clear prominence among non-restrictively licensed components, with Apache and BSD licenses far more represented in practice than GPL and LGPL. This differs from previous research, which indicated the reverse to be true.