

# **INTERACTIONS BETWEEN ECLIPSE FOUNDATION MEMBERS AND ECLIPSE PROJECTS**

by

Stephen James Anthony Lombardi

A thesis submitted to the Faculty of Graduate Studies and Research  
in partial fulfillment of the requirements for the degree of  
Master of Applied Science in Technology Innovation Management

Department of Systems and Computer Engineering

Carleton University

Ottawa, Canada, K1S 5B6

December 2008

© Copyright 2008 Stephen James Anthony Lombardi



Library and  
Archives Canada

Bibliothèque et  
Archives Canada

Published Heritage  
Branch

Direction du  
Patrimoine de l'édition

395 Wellington Street  
Ottawa ON K1A 0N4  
Canada

395, rue Wellington  
Ottawa ON K1A 0N4  
Canada

*Your file* *Votre référence*  
*ISBN: 978-0-494-47515-7*  
*Our file* *Notre référence*  
*ISBN: 978-0-494-47515-7*

**NOTICE:**

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

**AVIS:**

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protègent cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

---

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

  
**Canada**

The undersigned hereby recommend to  
the Faculty of Graduate Studies and Research  
acceptance of the thesis

**INTERACTIONS BETWEEN ECLIPSE FOUNDATION  
MEMBERS AND ECLIPSE PROJECTS**

Submitted by

Stephen James Anthony Lombardi

in partial fulfillment of the requirements for the degree of  
Master of Applied Science in Technology Innovation Management

---

V. Aitken, Chair

---

A.J. Bailetti, Thesis Supervisor

Department of Systems and Computer Engineering  
Carleton University  
Ottawa, Canada, K1S 5B6  
December 2008

## **ABSTRACT**

This research uses a data mining technique and information on company websites to classify the interactions between a sample of 163 Eclipse Foundation member organizations and Eclipse projects. The results suggest that Eclipse members: (i) use multiple Eclipse technologies to create offerings; (ii) use Eclipse as a platform for new application development or for offering complementary services; (iii) use open source business models to create offerings with Eclipse; and (iv) develop plug-ins for existing Eclipse applications. This research is relevant to the Eclipse Foundation and to researchers interested in: (i) processes for harvesting data on company websites and (ii) motivations for companies to be involved in open source projects.

## **ACKNOWLEDGEMENTS**

My deepest gratitude and thanks are extended to my thesis supervisor, Professor Antonio Bailetti. Without his support, direction and continued dedication this research would not have possible.

I would like to thank Professor John Callahan, Professor Thomas Kunz, Professor Steven Muegge, Professor Stoyan Tanev and Professor Michael Weiss for their expert knowledge and feedback which contributed greatly to the completion of this research.

I would also like to thank my fiancée Althea for her unending support, inspiration and encouragement throughout my graduate studies. Finally, I would like to thank my parents and my family for their motivation and support.

## **TABLE OF CONTENTS**

1. Introduction.....	1
1.1. Objective.....	1
1.2. Deliverables.....	2
1.3. Contributions.....	2
1.4. Relevance.....	2
1.5. Organization.....	3
2. Literature review.....	4
2.1. The Use of Websites for Corporate Communication.....	4
2.2. Business Ecosystems.....	12
2.3. Open Source.....	28
2.4. Lessons Learned.....	38
3. Research method.....	41
3.1. Unit of Analysis.....	41
3.2. Sample.....	41
3.3. Keyword Search Tool.....	41
3.4. Research Method.....	42
4. Results.....	46
4.1. Sample.....	46
4.2. Keywords.....	46
4.3. Descriptive Statistics.....	46

4.4. Factor Analysis .....	47
4.5. Summary of the Results Applied to the 163 Eclipse Members .....	52
5. Discussion of results .....	60
5.1. Interpretation of the Factors .....	60
5.2. Organization Clusters.....	61
5.3. Comparing Results with Member Websites .....	65
5.4. Developing New Applications and Interacting With Eclipse as Open Source .....	67
5.5. Interaction with Multiple Projects or Extension of Existing Applications ....	68
5.6. Comparison with Carbone’s (2007) Open Source Interactions .....	68
5.7. Distribution of Clusters by Eclipse Membership Type.....	70
6. Conclusions.....	73
6.1. Conclusions.....	73
6.2. Limitations .....	74
6.3. Future Research Opportunities .....	74
References.....	76

## LIST OF TABLES

Table 1: Summary of Business Ecosystems .....	17
Table 2: Descriptive statistics for the number of pages on an organization's website .....	47
Table 3: Commonalities for the variables used in the Factor Analysis with Alpha Factoring as the extraction method (sample size = 163).....	48
Table 4: Kaiser-Meyer-Olkin and Bartlett's test (sample size = 163).....	49
Table 5: Total Variance Explained by Extracted Factors (sample size = 163).....	50
Table 6: Factor Analysis Results (sample size = 163).....	51
Table 7: Identification and classification of the 163 Eclipse Member Organizations .....	52
Table 8: Distribution of Eclipse Members by Code .....	58
Table 9: Interpretation of the Extracted Factors .....	61
Table 10: Factor Loading by Clusters – A Comparison of Website data to Organization's Eclipse Interactions .....	66
Table 11: Comparison of Eclipse Interactions with Carbone's (2007) Open Source Interactions.....	69
Table 12: Distribution of Interaction Clusters by Eclipse Membership Type .....	71

## LIST OF FIGURES

Figure 1: Value network for the aerospace industry (Basole & Rouse, 2008) .....	24
Figure 2: Block diagram of the Eclipse platform.....	27
Figure 3: Recursive extensibility of the Eclipse platform .....	28
Figure 4: Stages in Carbone's open source maturity model (Carbone, 2007).....	33
Figure 5: The keyword search tool .....	42
Figure 6: Method for refining data matrix .....	44
Figure 7: Scree plot for Eclipse member organizations (sample size = 163). .....	50
Figure 8: Eclipse Member Organizations Grouped by the Codes Described in Table 8 (sample size = 163).....	62
Figure 9: Eclipse Member Organizations Grouped by the Codes Described in Table 8 (sample size = 163) (continued).....	63
Figure 10: Distribution of Organizations Grouped By Codes Described in Table 8 .....	64
Figure 11: Distribution of Associate Members by Interaction Clusters .....	72
Figure 12: Distribution of Solutions Members by Interaction Clusters.....	72
Figure 13: Distribution of Strategic Members by Interaction Clusters.....	72

## **LIST OF APPENDICES**

Appendix A: List of the 163 Eclipse Member Organizations in the Sample.....	81
Appendix B: Descriptive statistics for the variables representing keywords (sample size = 163).....	86
Appendix C: Factor Analysis Techniques .....	89

## **1. INTRODUCTION**

Much is known about the motivations, actions and value gained through an individual's interactions with open source projects (Hertel, Niedner & Herrmann, 2003; Lakhani & Wolf, 2005; Raymond, 2001). However, interactions between companies and open source projects have not been researched as extensively.

### **Eclipse and the Eclipse Foundation**

According to Eclipse.org (2008), "Eclipse is an open source community, whose projects are focused on building an open development platform comprised of extensible frameworks, tools and runtimes for building, deploying and managing software across the lifecycle. The Eclipse Foundation is a not-for-profit, member supported corporation that hosts the Eclipse projects and helps cultivate both an open source community and an ecosystem of complementary products and services."

#### **1.1. Objective**

The objective of this research is to use a data mining technique (Hicks, Libaers, Porter, & Schoeneck, 2006; Opoku 2005) and data on websites to classify companies that are members of the Eclipse Foundation by their products' interdependence with the Eclipse projects.

## **1.2. Deliverables**

The outcome of this research is the identification of the different ways Eclipse member companies refer to the interdependence between their products and Eclipse projects.

## **1.3. Contributions**

This research makes at least two contributions. First, the research provides a tool and an approach to identify the ways companies interact with open source foundations. The information can be used to:

- define and strengthen the customer value propositions that foundations offer their members
- segment members of open source foundations

The second contribution of this research is the development and introduction of an open source data mining tool known as Keyword Searcher Tool.

## **1.4. Relevance**

This research is relevant to the Eclipse Foundation and researchers. This research provides the Eclipse Foundation with a low cost approach to the identification of the ways that Eclipse member organizations use Eclipse projects. The information can be used by the Foundation to strengthen their value propositions to their existing members and to attract other members.

This research is also relevant to academics and researchers because it can rapidly gather a large set of data readily available on the Internet. Without the web search method, an expensive, long, tedious and resource intensive data collection process (interviews, case studies, etc.) needs to be conducted. Researchers are always looking for new techniques that can decrease the resources needed to collect data (Hicks, Libaers, Porter, & Schoeneck, 2006; Opoku,2005).

### **1.5. Organization**

This thesis has been organized into six chapters. The first chapter includes an introduction and overview of the research. The second chapter contains a review of the current literature pertaining to corporate websites, business ecosystems and open source software. In chapter three, the research strategy and method are described in detail. Chapter four covers the results of the data gathering and analysis. Chapter five contains a discussion of the results. Finally, chapter six covers the conclusions, limitations and future research opportunities.

## **2. LITERATURE REVIEW**

This chapter provides the review of the literature. It is organized into four sections.

Section 2.1 reviews the literature on the use of websites for corporate communication.

Section 2.2 reviews the literature on business ecosystems. Section 2.3 reviews the open source literature. Finally, section 2.4 provides lessons learned.

### **2.1. The Use of Websites for Corporate Communication**

Companies use their corporate websites to reach a diverse group of target audiences including investors, customers, employees, researchers and the media (Insch, 2008).

Companies use websites to advertise their corporate identity and their aspirations.

However, companies do not exploit the full potential of their websites as communications tools (Hicks et al., 2006; Pollach, 2005).

#### **2.1.1. Corporate Identity**

The vast majority of companies maintain a website. This has led to a wealth of research focusing on the content companies place on their website and on why and how it is reflective of the company's position in the marketplace.

Almost every firm's website has an "about us" page stating company goals, values and history. The underlying goal of the organizational website is to establish a corporate identity, not communicate with customers about products and services (Pollack, 2005).

Van Riel and Balmer (1997) state that “a corporate identity refers to an organization’s unique characteristics which are rooted in the behaviour of members of the organization.”

Corporate websites also describe their (i) products, (ii) actions in the marketplace, and (iii) organizational credentials.

Hicks et al. (2006) found that most websites are not used to interact with potential customers.

Most firms go beyond describing their current identity. They describe their potential and their aspirations (Hicks et al., 2006). Pollach (2005) found that even large and well established companies use their about page to make themselves look better than other firms in the industry. Extremely bold and ambitious claims were found to be the norm. Unverifiable claims were commonplace. For example, many organizations claimed dominance of the same market. Positive attributes were written in a passive voice to insinuate a general acceptance of them. Hicks et al. (2006) show that in defining their corporate identity, companies relate their current situation as well as their aspirations.

### **2.1.2. Challenges Faced by Corporate Websites**

Pollach (2005) lists usability, credibility and utility as major challenges faced by companies’ websites. Gustavsen and Bond (2003) report that interactivity is also a challenge faced by corporate websites. Interactivity adds a dimension that traditional media does not provide.

### **2.1.2.1. Site Usability**

At its most basic level, a website is a number of nodes and links. There tends to be more than one path through these nodes – users follow different links as they traverse the website. This makes website traversal rather complex when compared with traditional corporate literature. Pollach (2005) notes that in order to find the information they are looking for in a timely manner, users need content representation devices such as menus and textual organizers such as lists. She also notes that a user's perception of a website's content is positively correlated with the number of links a site has and the site's navigation structure.

Pollach's (2005) reports that most companies fail to provide an easily usable website specifically when it comes to within page navigation. Hicks et al.'s (2006) observe that most sites lack a "decent navigational scheme." Company websites lack contextual navigation aids such as "related pages" links (Pollack, 2005). Contextual navigation allows a company to guide users through its website in a specific path and results in increased page views by a user.

### **2.1.2.2. Message Credibility**

Pollach (2005) states that the general public tends to distrust corporate messages and are negatively biased towards companies. This leaves businesses facing an uphill battle when it comes to their website's credibility. She also notes that in addition to the public's distrust for corporate messages, the public also believes that information found on the

internet is not as trustworthy as information found in print. A website can increase its credibility by providing hyperlinks to websites of well-known, trusted organizations (such as government or non-governmental agencies) (Pollach, 2005). Also, brand loyalty affects a person's perception of message credibility when it comes to a corporate website – a person is more inclined to believe a message on the website of a company they are loyal to.

Pollach (2005) found that in her sample of companies' websites, a number of methods were used to increase credibility such as third-party evidence, testimonials and numerical data. However, when it comes to a company's description of itself and its position in the market exaggeration is common. Pollach (2005) warns that exaggeration can be dangerous because it may cause readers to leave a website. Readers will hardly be convinced by self-congratulatory statements with little information value. Sharp (2002) states that "users detest anything that seems like marketing fluff or overly hyped language ('marketese') and prefer factual information."

#### **2.1.2.3. Information Utility**

"A successful web site satisfies the goals of both the users and the provider of the website." (Pollach, 2005) For users, a corporate website is a collection of information about the company. For companies, the website should elicit a specific response from users. Sharp (2002) provides examples of positive responses such as: staying at the website for a certain amount of time, purchasing something from the site, visiting the company's brick and mortar location, or returning to the site at a later date.

Pollach (2005) found competing information regarding content as a determinant factor in the success of a website. In the case where a person visits a website with the intention of finding a specific piece of information than the quality of the website content will be paramount to providing a successful interaction. If however, a person is visiting the site simply to learn more about the company or browse the site than the navigation structure and other usability features will be more important than content.

Pollach's (2005) research found that for the most part, companies were not taking advantage of current technologies in order to create dialog and interaction between the website users and the company. By engaging users in a dialog a company can gather information which can be used to increase the sites usability and utility to the user.

#### **2.1.2.4. Interactivity**

Gustavsen and Bond (2003) decompose the interactivity of a website into six features: user control, personalization, responsiveness, connectedness, real time interaction and playfulness. They evaluated a number of corporate websites based on how many of these features were present. They found that on average site interactivity was low. On average, the sites in the sample had 43% of the interactivity features. Gustavsen and Bond (2003) found that the most interactive sites in their sample had high stick (staying at the website for a minimum duration) and return rates (coming back to the site at a later date).

### **2.1.3. Corporate Website Content Analysis**

The corporate website can be used as a valuable tool by researchers for collecting information. Data is much more readily available than using interviews or questionnaires and can be gathered quickly and efficiently using harvesting tools (Hicks et al., 2006).

#### **2.1.3.1. Automated vs. Manual Content Analysis**

A number of tools that can scour a website searching for relevant information exist today. These tools are referred to as spiders. Spiders can be used to find lists of sites based on a particular topic, can search for specific content within a predefined set of websites or choose search results based on any number of factors (Bauer & Scharl, 2000; Weare & Lin, 2000; Rennie & McCallum, 1999; Kosla & Bloxkeel, 2000; Stamatakis, Karkaletsis, Paliouras, Horlock, et al. 2003; Hicks et al., 2006).

Spiders can drastically reduce the amount of time needed to gather a large amount of data. The use of spiders has been considered a sound approach for data harvesting for almost ten years (McMillan, 2000). Bauer and Scharl (2000) also introduce a new framework for automated website content analysis involving three steps. In the first step, the web mirroring step – a snapshot of the websites to be analyzed is taken by saving all of the website's data. The second step involves the extraction of the classification criteria. This is done using a customized application that extracts variables in three groups: content, interaction elements and navigational elements by examining the web pages' HTML tags. The final step is the analysis step in which a combination of the following methods are used in order to classify the sites: manual classification, textual analysis, statistical clustering and neural network based classification.

Perry and Bodkin (2000) examined self promotion by Fortune 100 companies on their websites. McMillan (2000) surveyed 19 different studies using web content analysis. Pollach (2005) examined how companies present themselves on their websites. Porter, Schoeneck, Frey, Hicks, and Libaers (2007) described a method for collecting technical intelligence and Opoku (2005) investigated the most innovative business schools.

### **2.1.3.2. Web Content Analysis Methodologies**

Content analysis has been in existence since as early as the 1600s but became a well known term in the early 1960's (Krippendorff, 2004). Since then the methodology for content analysis has been solidified and an enormous amount of literature has been analyzed using this methodology. Web content analysis follows similar methodologies depending on the tool used. This section will describe two different methodologies for web content analysis: McMillan's (2000) take on traditional content analysis applied to the web and Hicks' (2006) web search method for content analysis.

The methodology laid out in McMillan (2000) is made up of five steps. The first step is formulating hypotheses regarding the websites to examine. The second step is to select a sample of websites. McMillan notes that popular ways of selecting a sample include using an online list of websites for a particular category or using web searches to identify a number of related websites. The third step in the process is to define categories and collect data. In traditional content analysis the timeframe is used to describe the publishing date of the material. In web content analysis however, the time period is the

timeframe in which the data was gathered from the Internet. The dynamic nature of websites necessitates a rapid collection of data so that there is consistency across websites (McMillan, 2000). Data is collected by coders who traverse the site based on the categories defined (i.e., for a set amount of time or until a certain list of categories have been found on the site). Results from these coders need to be verified so that the research can be reproduced with a certain amount of predictability. The fourth step is testing the reliability of the coding. This can be done by analyzing data from cross coded websites (sites analyzed by more than one coder). The fifth and final step is to analyze the data collected by the coders.

Hicks et al. (2006) method for website content analysis is quite different in that all of the data is collected by querying search indexes for material collected using spiders. This method allows for much more rapid content analysis since there is no need for coders. This also eliminates some of the bias and inconsistency introduced by using multiple coders. The web search method is used to group websites into clusters based on a set of keywords used on the sites. The method has three main steps. The first step is keyword development in which a list of keywords is created that will be used to define the clusters of websites. The list of websites is also selected using similar methodologies as those described in step one of McMillan's (2000) paper. The second step is the website search in which each of the websites is searched for each of the keywords and the number of hits (pages within the website containing the keyword) is recorded. The final step is a factor analysis of the results in order to sort the list of websites into clusters (factors).

## **2.2. Business Ecosystems**

Ecosystems have been studied for as long as biologists have been studying the earth we live in. Recently, however, the analogy of the business ecosystem has been introduced to describe a number of organizations' interactions. This section examines business ecosystems and describes the Eclipse ecosystem.

### **2.2.1. Definition of a Business Ecosystem**

In Peltoniemi and Vuori (2004) the varying definitions of a business ecosystem are summarized and commented on and based on these, a composite definition is formed. This section will outline each of the different definitions of a business ecosystem in their (2004) survey.

#### **2.2.1.1. Moore's Ecosystem**

Moore has defined a business ecosystem in two slightly different ways. Peltoniemi and Vuori include both definitions. In his first definition, Moore (1996) defines a business ecosystem as "an economic community supported by a foundation of interacting organizations and individuals – the organisms of the business world." The ecosystem includes customers, lead producers, competitors and other stakeholders. Ecosystem leaders also known as the "keystone species" have a lot of influence in the evolution of the ecosystem. In his second definition, Moore (1998) defines a business ecosystem as an "extended system of mutually supportive organizations; communities of customers, suppliers, lead producers, and other stakeholders, financing, trade associations, standard

bodies, labor unions, governmental and quasi governmental institutions, and other interested parties. These communities come together in a partially intentional, highly self-organizing, and even somewhat accidental manner.” Peltoniemi and Vuori (2004) note the interesting difference between the two definitions – the first focusing on the interaction between companies within the ecosystem and the second focusing on self-organization of the ecosystem and decentralized decision making by the parties involved.

Peltoniemi and Vuori (2004) examine Moore’s definition of stages of development. The first stage, also known as the birth stage, requires that ecosystem members more than satisfy customers in order to ensure ecosystem health and future growth. The second stage, known as expansion, is when the ecosystem begins to grow and scaling potential is tested to ensure rapid expansion is possible. In the leadership stage, high-profitability is achieved along with stability and ecosystem health. In the fourth and last stage the ecosystem is renewed or collapses due to outside threats.

#### **2.2.1.2. Iansiti and Levien’s Ecosystem**

Iansiti and Levien (2004) use the analogy of a biological ecosystem to describe complex business networks. They use the analogy to garner a better understanding of these systems:

“Like business networks, biological ecosystems are characterized by a large number of loosely interconnected participants who depend on each other for their mutual effectiveness and survival. And like business network participants, biological species in ecosystems share their fate with each other.” (Peltoniemi and Vuori, 2004)

Individuals in the ecosystem cannot prosper or suffer alone. For any member to be successful, the ecosystem as a whole must thrive. It is important to note that swings from a healthy to an unhealthy ecosystem can take place very rapidly and once the delicate balance within the ecosystem is disturbed it is hard to regain. Once sustainability has been reached in Iansiti and Levien's (2004) ecosystem, individual organizations may be swapped in and out of the ecosystem with no detrimental effect to the ecosystem. This modularity allows the ecosystem to self-govern its participants much like the way animals are banished in a biological ecosystem if they are not performing their duties. This helps to ensure the health of the ecosystem (Iansiti & Levien, 2004).

Iansiti and Levien (2004) identify three critical success factors for a healthy ecosystem: productivity, robustness and niche creation. Productivity is important, especially early on in the life of the ecosystem in order to reach a critical mass. Without productive members, there will not be enough value generated in the ecosystem to support all of the members.

Iansiti and Levien (2004) define robustness as the ability to draw a competitive advantage from a number of sources and having the ability to redefine offerings based on market changes. Robustness is important for each of the members of the ecosystem in order to ensure the ecosystem itself is robust. Robustness is not as important in later stages of the ecosystem's life cycle when individual organizations can be swapped in and out but it is critical when the number of participants in the ecosystem is small and each participant plays a vital role.

According to Iansiti and Levien's (2004), an organization in an ecosystem takes on one of four different roles: keystones, niche players, dominators, or hub landlords.

Keystone players are tasked with the responsibility of maintaining ecosystem health. They must put more value into the ecosystem than they draw from it. Keystones have the most power in the ecosystem and exert the most influence over it even though they make up the smallest number of organizations within the ecosystem. Keystones are not necessarily the largest organizations in an ecosystem, but they are a central hub through which many of the interactions take place.

Niche players tend to make up the largest portion of the ecosystem and have the least influence and power over it.

Dominators and hub landlords control portions of the ecosystem. They take much more value from the ecosystem than they put into it. Iansiti and Levien (2004) state that dominators can lead to ecosystem instability if too many arise and are not concerned with the health of the ecosystem or even their own long term health, but rather short term profits.

#### **2.2.1.3. Power and Jerjian's Ecosystem**

Peltoniemi and Vuori (2004) summarize the definition of Power and Jerjian's (2001) ecosystem as "a system of websites occupying the World Wide Web, together with those

aspects of the real world with which they interact. It is a physical community considered together with the non-living factors of its environment as a unit.” In order to turn a biological ecosystem into this model, the organisms are substituted with websites, and the habitat is substituted with the World Wide Web infrastructure. They define four stakeholders to an organization which are: communities of shareholders, employees, businesses and customers. The authors note that of the ecosystems examined, Power and Jerjian’s (2001) ecosystem definition is the most abstract. Peltoniemi and Vuori (2004) state that the theory

“would be better in line with [Power and Jerjian’s] other statements to say that many businesses are so highly dependent on WWW that it is their virtual habitat where web sites are their expressions”

#### **2.2.1.4. Lewin and Regine’s Ecosystem**

Peltoniemi and Vuori (2004) summarize Lewin and Regine’s (1999) ecosystem definition as “a network of companies each occupying a place on its own landscape of possibilities, and each landscape being coupled to many others: those of competitors, collaborators, and complementors.” This ecosystem mimics that of a biological ecosystem because, due to the interconnectedness of the organizations, changes to individual organizations dramatically affect other landscapes. They recommend that organizations in complex ecosystems base their strategy around co-evolution concepts.

#### **2.2.1.5. The Multiproject Software Ecosystem**

Scacchi, Feller, Fitzgerald, Hissam and Lakhani (2006) describe a multi-project software ecosystem in which the ongoing development and evolution of the open source software

system leads to changes, vulnerabilities and strengths in the projects linked to it.

Interdependencies between open source projects become most obvious when source code is actually shared between the projects.

### 2.2.1.6. Peltoniemi and Vuori's Ecosystem

Peltoniemi and Vuori (2004) contribute their own definition of a business ecosystem, based on all of the other definitions. The resultant definition is as follows:

“We consider a business ecosystem to be a dynamic structure which consists of an interconnected population of organizations. These organizations can be small firms, large corporations, universities, research centers, public sector organizations, and other parties which influence the system...Business ecosystems develop through self-organization, emergence and co-evolution, which help it to acquire adaptability. In a business ecosystem there is both competition and cooperation present simultaneously.”

**Table 1: Summary of Business Ecosystems**

Author	Main Points
Moore (1996, 1998)	"Extended system of mutually supportive organizations; communities of customers, suppliers, lead producers, and other stakeholders... These communities come together in a partially intentional, highly self-organizing, and even somewhat accidental manner."
Iansiti and Levien (2004)	<p>“Like business networks, biological ecosystems are characterized by a large number of loosely interconnected participants who depend on each other for their mutual effectiveness and survival. And like business network participants, biological species in ecosystems share their fate with each other.”</p> <p>Productivity, robustness and niche creation are key ecosystem success factors.</p>
Power and Jerjian (2001)	“a system of websites occupying the world wide web, together with those aspects of the real world with which they interact. It is a physical community considered together with the non-living factors of its environment as a unit.”

Lewin and Regine (1999)	“a network of companies each occupying a place on its own landscape of possibilities, and each landscape being coupled to many others: those of competitors, collaborators, and complementors.”
Multiproject Software Ecosystem (Scacchi et al., 2006)	The ongoing development and evolution of the open source software system leads to changes, vulnerabilities and strengths in the projects linked to it.
Peltoniemi and Vuori (2004)	"A dynamic structure which consists of an interconnected population of organizations... Business ecosystems develop through self-organization, emergence and co-evolution, which help it to acquire adaptability. In a business ecosystem there is both competition and cooperation present simultaneously."

### 2.2.2. The Eclipse Ecosystem

“Eclipse is an open source community, whose projects are focused on building an open development platform comprised of extensible frameworks, tools and runtimes for building, deploying and managing software across the lifecycle. The Eclipse Foundation is a not-for-profit, member supported corporation that hosts the Eclipse projects and helps cultivate both an open source community and an ecosystem of complementary products and services.” (Eclipse Foundation, 2008)

The Eclipse Foundation is unique in that it actively assists in the marketing and promotion of Eclipse projects and products and services offered by its members (Eclipse Foundation, 2008). The Eclipse Foundation (2008) believes that open source projects and commercial products based on Eclipse as well as publications, training and consultation and service providers are all important facets of the ecosystem and are important to the health of the ecosystem as a whole. The Eclipse Foundation (2008) helps maintain and grow the Eclipse ecosystem through actions such as community events and conferences, marketing events and advertising, meetings and other activities that promote Eclipse.

## **A Brief History of Eclipse**

In 2001 after the acquisition of a small company in Ottawa with a software development tool, IBM decided to “run Eclipse as the marriage of an open source project community and an industry consortium” the result of which has “been a major source of [the Eclipse] community's enduring strength and growth” (Smith & Milinkovich, 2007). As further versions of Eclipse were released in 2002 and 2003 the consortium surrounding Eclipse began expanding rapidly. In 2004, Eclipse 3.0 was released based on open standards and the Eclipse Foundation was created (Smith & Milinkovich, 2007; Eclipse Foundation, 2008).

### **2.2.2.1. Eclipse Foundation Members**

At the time of writing there are five Eclipse membership types available: Associate Members, Solutions Members, Enterprise Members, Strategic Members, and Committer Members.

#### **Associate Members**

Associate membership is free and open to non-for-profit organizations, standards bodies, universities, research institutes and media organizations. Membership is also open to for-profits with annual dues. These members do not sit on the board of directors but are able to submit requirements and participate fully in reviews and membership meetings as stated in Eclipse Foundation (2008c).

**Solutions Members (previously Add-In Providers)**

Solutions members must cross-promote the Eclipse ecosystem with a public statement of their involvement in Eclipse and they must offer a commercial product based on or using an Eclipse project within one year of joining the membership (Eclipse Foundation, 2008c). Solutions members pay annual dues commensurate with their annual revenue as is laid out by the Eclipse Foundation (2008c).

**Enterprise Members**

At this time, the enterprise membership class has not been formally launched but has been defined and will be launched in the near future. An enterprise member is a large organization that depends on Eclipse technology in order to produce its market offerings (Eclipse Foundation, 2008c). Enterprise members actively participate in the planning and development of the Eclipse Foundation and the Eclipse ecosystem. Enterprise members pay annual dues as is laid out by the Eclipse Foundation (2008c).

**Strategic Members**

Strategic membership gives an organization a place on the Eclipse board of directors which allows them to directly influence the direction of Eclipse; members also have a seat on the Eclipse requirements council which gives them an influence in the direction of Eclipse technology (Eclipse Foundation, 2008c). There are two kinds of strategic members: Strategic developers and Strategic consumers. Strategic developers make large contributions to the Eclipse codebase through hiring full-time developers. Strategic

consumers are major users of Eclipse technology. Strategic members pay annual dues set by the Eclipse Foundation (2008c).

### **Committer Members**

Committer members make contributions to the Eclipse codebase as part of a larger organization or as individual members. There are no dues for committers. Committers do not attend membership meetings but can vote for committer representation on the Eclipse board of directors (Eclipse Foundation, 2008c).

#### **2.2.2.2. Success of the Eclipse Ecosystem**

The Eclipse ecosystem has thrived since the release of Eclipse 1.0 in 2001 and the founding of the Eclipse Foundation in 2004. The Eclipse Foundation has over 170 member organizations and thousands of code contributors worldwide (Eclipse Foundation, 2008). Smith and Milinkovich (2007) believe this success is due to a number of key factors including:

- a vendor neutral governance model
- organizational diversity
- a common technical architecture
- three factors identified by Iansiti and Levien (2004): productivity, robustness and the ability to create niches

### **Vendor Neutral Community**

The Eclipse Foundation was established as and continues to be an independent, not-for-profit corporation to help administrate and guide the Eclipse ecosystem. The foundation has allowed for an independent, vendor neutral community to establish itself around the Eclipse projects. This community has seen a rapid expansion with a number of large industry players signing on as strategic developer members. (Eclipse Foundation, 2008)

### **Diversity in the Ecosystem**

The Eclipse ecosystem has over one hundred and seventy contributing members that are developing new products and tools based on Eclipse (Eclipse Foundation, 2008). This kind of diversity ensures ecosystem health because there are so many different contributing entities that the success or failure of the ecosystem as a whole cannot be determined by any one of the contributing entity's success or failure (Smith & Milinkovich 2007).

### **Common Technical Architecture**

Eclipse is built on the OSGI specification (OSGI, 2008). OSGI is an open, extendable framework that allows the provision of plug-ins to add functions or features to an existing application. This flexibility allows any number of applications to be built on top of Eclipse ranging from development tools to email clients to application servers and everything in between (Smith & Milinkovich, 2007). This flexibility allows Eclipse to go beyond its roots as a development tool in any number of directions increasing the ecosystem's diversity as it does.

**Productivity**

Productivity here is defined as the Eclipse ecosystem's ability to create economic value. Smith and Milinkovich (2007) state that as more value is being created within the Eclipse ecosystem, more jobs are becoming available which leads to more value creation. This loop of value creation also entices new members to join the fray and begin producing products based on Eclipse (Smith & Milinkovich, 2007). New entrants not only provide more value within the ecosystem, they also increase the ecosystem's diversity.

**Robustness**

Smith and Milinkovich (2007) identify two main competitors to the Eclipse community: Microsoft and Sun. This is due to the fact that Eclipse's main franchise is its Java integrated development environment (IDE) and both Microsoft and Sun provide products with similar functionality (Smith & Milinkovich, 2007). Although Sun has made large investments in their offerings, the Eclipse market share and download rates continue to increase (Smith & Milinkovich, 2007). According to Smith and Milinkovich (2007) Eclipse's growth in the face of strong competition is a good indicator of the robustness of the Eclipse ecosystem.

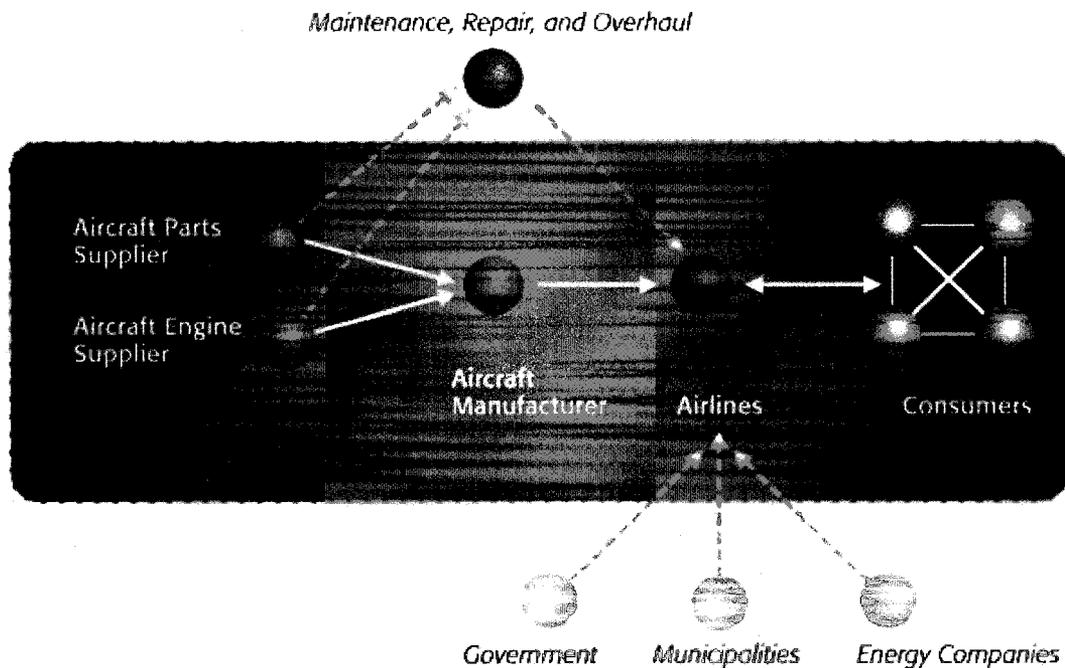
**Niche Creation**

The Eclipse platform has been very successful in enabling the creation of niches around its offerings (Smith & Milinkovich, 2007). Smith and Milinkovich (2007) believe that the

extensibility of the OSGI-based frameworks makes Eclipse adaptable to new technologies and allows for rapid niche development.

### 2.2.2.3. Value Networks

Christensen and Rosenbloom (1995) define a value network as a nested network of producers and markets through which the tradable architected components at each level are made and sold to integrators at the next higher level in the system. Christensen (2004) adds to that explaining that “the context within which a firm identifies and responds to customers' needs, solves problems, procures input, reacts to competitors, and strives for profit—is central to [a firm's value network].”



**Figure 1: Value network for the aerospace industry (Basole & Rouse, 2008)**

Christensen (2003) proposes five success factors for innovative firms in a value network.

These factors are:

1. The value network a firm resides in has a large impact on that firm's ability to overcome both organizational and technical challenges as it strives for innovation.
2. In order to be successful, a firm must address the needs of known actors within their network.
3. Established firms must not ignore new technologies that are maturing at a rate that is faster than the technologies they currently lose no matter how immature they are at the time.
4. Entrant firms have an advantage when it comes to these disruptive technologies and established firms must adopt quickly to avoid losing ground later.
5. This advantage is due to the relative flexibility with which entrant firms can change their strategies (their agility).

#### **2.2.2.4. The Eclipse Ecosystem as a Value Network**

Christensen and Anthony (2004) have identified six categories used to evaluate a value network: suppliers, customers, retailers, distributors, partners and ancillary industry players.

McGregor (2008) uses these categories to analyze the Eclipse Foundation's value network. The results are as follows:

**Upstream suppliers:** A number of other open source projects, as well as individual developers and employees hired by companies in order to develop code for the Eclipse projects.

**Downstream customers:** Software development projects built on Eclipse including commercial software vendors.

**Retailers:** Eclipse is not sold per se, however product integrators provide tools based on Eclipse projects. Sale of these products is allowed by the Eclipse Public License (see Eclipse Foundation, 2008b)

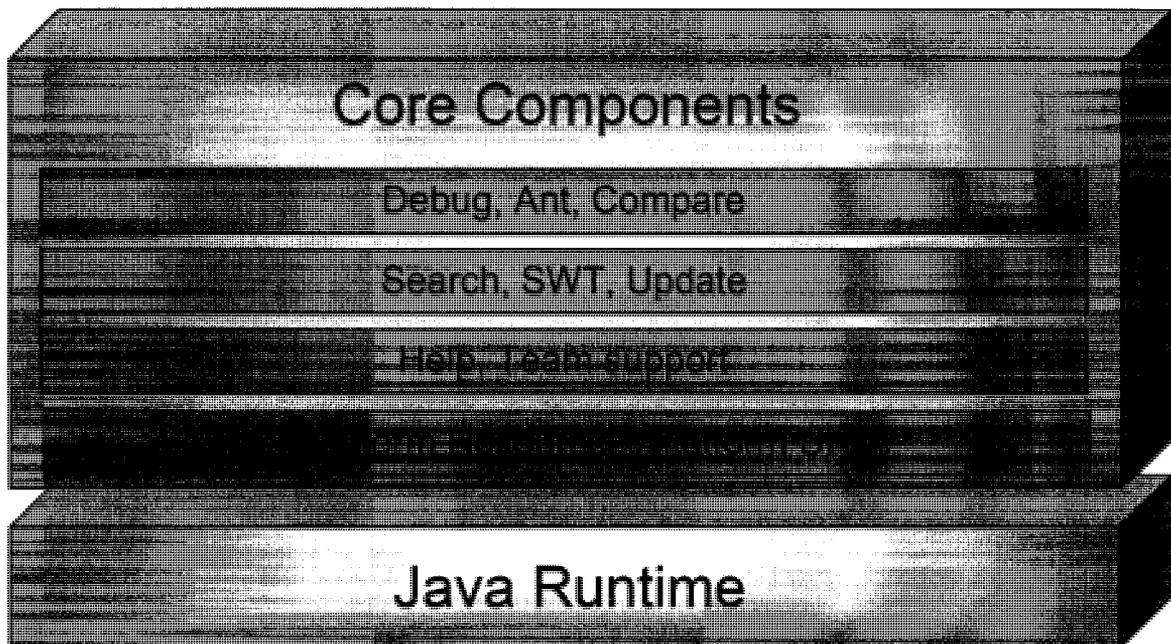
**Distributors:** Eclipse is distributed by a number of mirror sites including research universities, private companies and third party software download sites.

**Partners:** The Eclipse Foundation (2008) has over one hundred and seventy member organizations which are deemed to be partners. The members support the development of Eclipse through administrative positions and financial support.

**Ancillary industry players:** Includes users who download Eclipse, users who participate in forums and users who open defect reports. Future suppliers come from this group.

### 2.2.3. The Eclipse Platform

Built on OSGI, Eclipse can be extended in order to enhance existing applications or create entirely new ones (Eclipse Foundation, 2008d). The platform consists of a set of core plug-ins running in a Java runtime environment. Figure 2 provides a block diagram of the Eclipse platform.

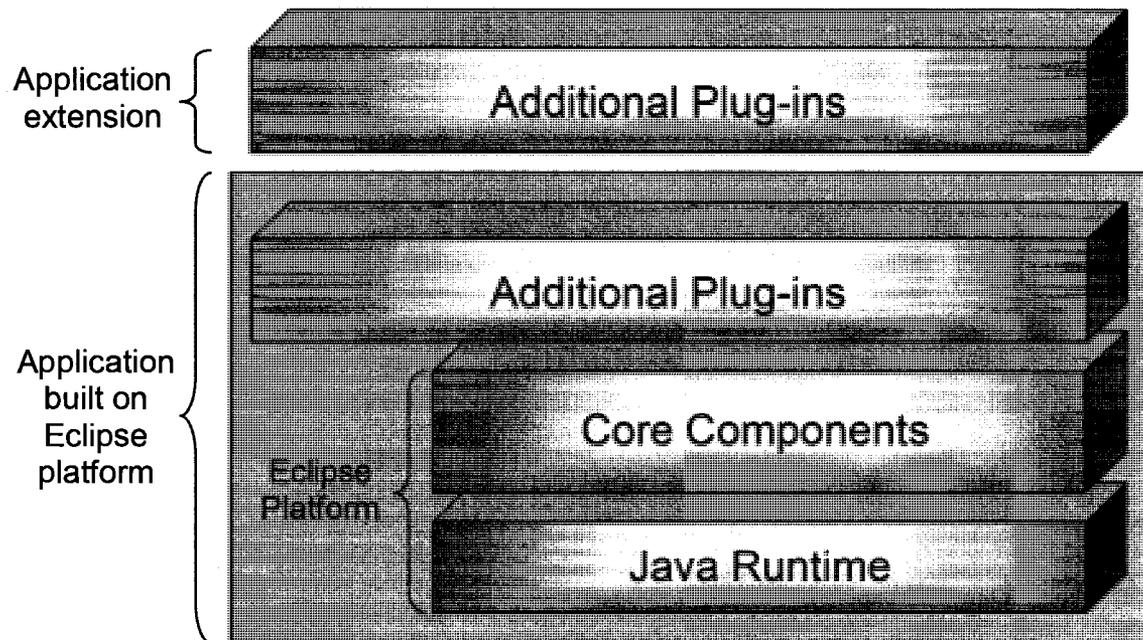


**Figure 2: Block diagram of the Eclipse platform**

The functionality of the Eclipse platform is extended by adding additional plug-ins on top of the core components. Small features can be added in a single plug-in but more complex additions often consist of multiple plug-ins.

Entirely new applications can be developed by contributing new plug-ins. These new applications can be extended by contributing new plug-ins. New application can be built

on top of these new applications in a recursive manner. Figure 3 illustrates the extensibility of the Eclipse platform:



**Figure 3: Recursive extensibility of the Eclipse platform**

## 2.3. Open Source

Large corporations are using open source to generate income (Koenig, 2004) and researchers and industry players alike praise the many benefits of open source (Lussier, 2004; Raymond, 1999).

### 2.3.1. The Benefits and Drawbacks of Open Source Adoption

Open source can be beneficial in some situations and totally inapplicable in others. This section will investigate the benefits, constraints and caveats in adopting open source software.

### **2.3.1.1. Driving Factors and Benefits of Open Source**

There are many reasons why organizations adopt open source software (Bauer & Pizka, 2003; Merilinna & Martinlassi, 2006; Polancic, Horvat & Rozman, 2004; Wang & Wang, 2001). These reasons include: the demand for higher quality software at the same or lower cost, the need to rapidly bring applications to market, and customer demand for highly upgradeable software.

The open source community can provide the tools necessary to streamline product development at a reduced cost. Bauer and Pizka (2003), Fitzgerald (2006), Lussier (2004), Merilinna and Martinlassi (2006), and Polancic et al. (2004) provide data and anecdotal evidence in support of a number of benefits of open source including the availability of free tools, development frameworks and delivery platforms, applications with an active developer community and access to rapid defect resolution and feature additions, highly customizable tooling which can be tailored to suit a specific need, industry tested and trusted server software such as Apache and MySQL, development frameworks and languages that have re-use and upgradeability as prime concerns, and tools that allow new developers to become productive faster.

### **2.3.1.2. Limitations of Open Source**

Although open source development processes have come a long way since their early days of unstructured hacking (Fitzgerald, 2006), there are still a number of factors an

organization must consider before adopting an open source project for use or development (Merilinna & Martinlassi, 2006).

The availability of technical support is an important factor to consider since, for many open source applications, commercial grade technical support is not available. This could lead to a difficult situation should problems with the code arise.

The open source project's compatibility with open standards is also an important factor to consider. Standards-based projects will be easier to integrate and easier to extend in the future. This leads to the next factor which is the customizability and extensibility of the code. It is likely that the code does not provide exactly what an organization needs; however if the project has been developed in such a way that it can be extended and customized it can still provide a viable solution.

If the open source being adopted is for use in an enterprise application, it must have a high level of robustness and reliability. Usually this means it has had extensive testing and use in other, similar environments. In this same area, development complexity of the project should be considered if customization or extension is expected. Some projects are more complex than others and require special skills.

Although open source is technically "free" there are a number of costs associated with its adoption. In some cases the cost of customization and legal costs can be more than purchasing a license for a similar piece of proprietary software. An organization should

ensure it has a realistic budget to support code adoption. With open source comes open source licensing. Organizations must understand the terms of the licenses they are using and understand their implication on the company's business practices. Legal advice is often a part of the budget for adopting open source.

Although there are limitations and caveats to adopting open source, many of these exist in the proprietary software market as well. Since every organization is different, its needs will be different and decisions to use an open source project should be made on a case by case basis (Merilinna & Martinlassi, 2006).

### **2.3.2. Carbone's Open Source Interactions**

Carbone (2006, 2007) outlines five types of open source interactions for organizations. He also examines critical success factors and challenges for achieving open source maturity. This section will cover Carbone's open source interactions and the seven critical success factors he identifies for open source projects.

#### **2.3.2.1. Open Source Success Factors**

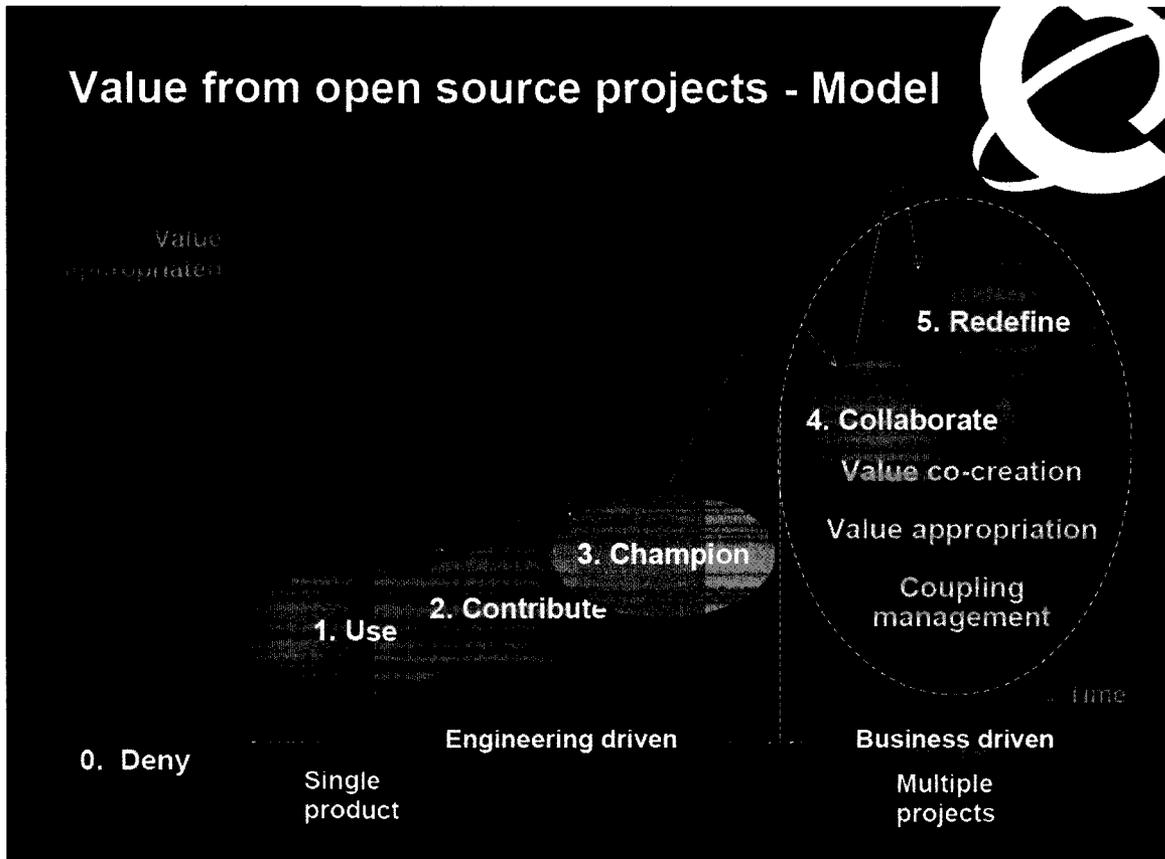
Carbone (2007) defines seven factors critical to the success of a company using open source:

- ability to appropriate the value created by the open source ecosystem (this factor outweighs the cost savings associated with open source)
- ability to compete; a company can't win just because the software it uses is free

- ability to influence the open source community
- ability to create good experiences for customers who use open source
- ability to attain a leadership position and enjoy the recognition that the position affords
- ability to co-create value with other organizations
- ability to ensure that open source is applicable to the task the organization's offering performs.

#### **2.3.2.2. Carbone's (2007) Open Source Interactions**

Carbone (2006, 2007) defines 5 open source interactions for organizations. In the user interaction, an organization's employees use open source software in order to decrease cost or development time or increase product quality. In the contributor interaction, managers contribute resources and source code to an open source project. This is beneficial to both the organization and the open source community as they both receive the benefit of a more complete, stable product. In the champion interaction, the organization



**Figure 4: Interactions in Carbone's open source model (Carbone, 2007)**

provides a well-known and experienced manager to act as a leader for the software project. This interaction provides the organization with a competitive advantage in their market as a result of their interactions with open source. In the collaborate interaction, an organization defines their position in the ecosystem around the use of open source. Using this interaction, the organization acts across a number of projects in order to change the competitive environment to one that favours them. Finally, in the redefine interaction, an organization redefines its offers based on the needs of the ecosystem. The organization gains a competitive advantage by harnessing changes in multiple ecosystems.

### **2.3.3. Open Source Business Models**

A number of organizations drive revenue through interactions with open source and open source ecosystems. A number of large industry players such as IBM, Oracle, Novell and HP have all pledged support for open source and some of them are making more than a billion dollars in annual revenue from open source business models. Koenig (2004) has identified seven popular business models anchored on open source and this section reviews these models.

#### **2.3.3.1. Optimization Strategy**

Organizations that use the optimization strategy use modular platforms to their advantage. They realize that commoditized layers in the platform are not worth the development and maintenance costs incurred in producing them (Raymond, 1999). Rather than develop these commodities in-house, optimizers use open source frameworks and deliver optimized products based on them. This decreases production cost and time to market while lowering product cost and increasing market share (Koenig, 2004). An example of an optimizing organization is Oracle and their Real Application Cluster (RAC) built to run on commodity Linux servers.

#### **2.3.3.2. Dual Licensing Strategy**

Organizations that use the dual licensing strategy offer one or more products with two licenses. Products licensed under an open source license typically have a reduced feature

set or have usage restrictions, commercial or otherwise (Koenig, 2004). Koenig (2004) uses MySQL as an example of a dual licensor following the free for non-commercial use, pay for commercial use template.

#### **2.3.3.3. Support Strategy**

In the support, consulting, patronage and hosting strategy organizations act as service providers rather than product providers. They offer services that complement open source offerings in order to create revenue (Koenig, 2004). In the support strategy there are two main streams. The first involves offering a product as open source and giving users the option of subscription based support. The second is to provide support for one or more open source projects that were not developed by your organization (Koenig, 2004). JBoss is one example of an organization that drives the production of and offers commercial support for an open source project (Zachary, 2006). Although Red Hat does drive the production of Red Hat Linux, they also offer support for a number of other open source projects and fall into the latter category of support strategists (Zachary, 2006).

#### **2.3.3.4. Consulting Strategy**

Using the consulting strategy in order to drive revenue through open source is a growing trend as enterprise-based revenue is increasingly coming from services (Koenig, 2004). The consulting strategy involves the creation of a total solution around open source that includes hardware, software and support. Red Hat is one company that uses the consulting strategy, with Koenig (2004) noting that an average of only 4% of the cost of a

Red Hat Linux solution is in Red Hat fees. Certification programs such as those offered by Sun and Novell are also consulting strategy activities (Koenig, 2004). These certifications provide two benefits: (i) they create revenue from certification and (ii) raise the reputation of the product.

#### **2.3.3.5. Patronage Strategy**

In the patronage strategy, an organization provides resources, support and publicity to an open source project. This allows an organization to drive the creation of standards around their championed project and it increases the use of the project's open source code which leads to more customers and drives sales of the organizations commercial offerings (Koenig, 2004). IBM has successfully employed the patronage strategy with Eclipse. IBM's initial investment of forty million dollars worth of intellectual property has led to an entire generation of software developers being introduced to Eclipse during their university training. This kind of familiarity with Eclipse products drives sales of IBM's commercial Eclipse-based offerings such as the Rational suite of development tools (Koenig, 2004).

#### **2.3.3.6. Hosting Strategy**

The hosting strategy involves organizations offering products as subscription based services. Examples of these services include: Google AdWords, Amazon S3 and E-Trade. Typically these services run on Linux servers utilizing open source application stacks in order to keep implementation costs to a minimum while providing services that would

cost more to the end user if stand alone licenses were purchased to perform the same tasks (Koenig, 2004).

#### **2.3.3.7. Embedded Strategy**

Organizations that produce embedded systems are able to leverage open source software and operating systems in order to provide a robust and extensible platform with minimal capital output up front (Koenig, 2004). An example of this is a product such as Tivo, which runs on Linux software. Tivo's revenue is driven from the cost of hardware and through monthly service fees (Tivo also uses the hosting strategy). Linux provides a number of distinct benefits in the embedded market including "stability, small foot-print, and networking" (Koenig, 2004).

#### **2.3.3.8. Combined Models**

When an organization decides to use open source as a way to create value and drive revenue they have a number of different business models to choose from. Often an organization chooses to use a combination of the above business models as well as other models not mentioned here in order to create total solutions for customers and to create the most value for their organization (Koenig, 2004).

## **2.4. Lessons Learned**

This section outlines the lessons learned from the literature reviewed.

**The “About us” pages in corporate websites are used to communicate image but the websites are not used to their full potential as mediums for public communication**

Corporate websites successfully communicate their values, goals and vision using their “about us” page, but more could be done in this area, specifically with regards to building trust with the public (Insch, 2008; Sullivan, 1999)

**Co-operative interactions within an ecosystem are important to ecosystem health**

A significant amount of time and effort has been spent defining business ecosystems and investigating how they work (Peltoniemi & Vuori, 2004). Iansiti and Levien (2004) believe that productivity, robustness and the ability to create niches are critical factors in the health of the ecosystem. Companies cooperate to create value and compete to appropriate the value they create.

**The Eclipse Foundation defines five types of members**

The Eclipse Foundation (2008c) defines five types of members. Four of the membership types are for organizations; one is reserved for individual members.

**A company profits from open source projects in different ways**

A number of business models have emerged that generate revenue through interactions with open source. These business models are both product and service driven (Koenig,

2004) and focus on separating commodity from competitive advantage (Raymond, 1999). A number of large companies are using these models to drive open source revenues of over one billion dollars annually (Koenig, 2004).

Carbone (2006, 2007) defines five levels of open source maturity based on the reasons why an organization interacts with an open source project.

Open source maturity involves the ability to harness value across multiple ecosystems within the open source community and exert influence within them to create a favourable environment (Carbone, 2007). In order to achieve this sort of maturity organizations must be able to appropriate the value created by the open source ecosystem, they must understand that the open source software ecosystem is competitive; they have to be able to actively manage their open source community and they must provide customers with a good open source experience.

**Data mining techniques are well suited to investigating open source projects.**

Because most open source projects have websites, data mining techniques can be particularly effective for open source research. An open source project's website is often a wealth of information including press releases, product information, release information and dates, feature requests and defect tracking. Polancic, Horvat and Rozman (2004) take advantage of this availability of information online in developing their assessment technique of open source software.

**There is a limited amount of research investigating organizations' interactions with open source projects.**

There has been a lot of research investigating interactions between individuals and open source, open source projects and their end users and how organizations use open source to create value (Hertel, Niedner & Herrmann, 2003; Lakhani & Wolf, 2005; Raymond, 2001). However, there is little in the literature that examines how organizations interact with open source projects.

### **3. RESEARCH METHOD**

This chapter describes the research design and method. Section 3.1 identifies the unit of analysis. Section 3.2 describes the sample used in this research. Section 3.3 describes the Web Search method. Section 3.4 contains the research method.

#### **3.1. Unit of Analysis**

The unit of analysis is the website of an organization that was a member of the Eclipse Foundation as of October 5, 2008.

#### **3.2. Sample**

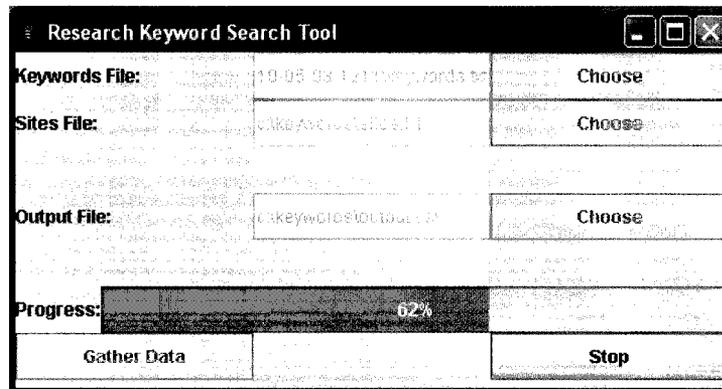
The following criteria were used to select the sample:

- The website belongs to an organization that was listed as a member of the Eclipse Foundation in Eclipse.org membership list as of October 5, 2008.
- The website had at least ten pages.
- The Eclipse member is not a media or educational organization

#### **3.3. Keyword Search Tool**

The web search method used by Hicks et al. (2006) and Opoku (2005) will be used to generate the deliverables. This method requires many internet searches to be run in order to collect data for analysis. In order to speed the data collection, a keyword search tool has been developed as part of the contribution of this research.

Figure 5 illustrates the interface of the keyword search tool that was used in this research.



**Figure 5: The keyword search tool**

The keyword search tool was developed in Java and uses the Google (2008b) AJAX application programmer interface in order to make web searches. The tool was designed to make multiple searches concurrently for speed. Several thousand queries can be completed in just a few minutes, saving the researcher many hours. Two text files, one containing a list of websites and the other containing a list of keywords are provided to the tool. The tool searches each website for each of the keywords and records the occurrences. The tool normalizes data by converting the number of occurrences into a percentage of pages on the site on which the keyword occurs. This normalized data is then saved for analysis.

### **3.4. Research Method**

The research method had five key steps:

1. Define the list of keywords that will be used in the data mining tool

2. Use the keyword search tool to collect data from the websites of the organizations that are members of the Eclipse Foundation
3. Refine the raw data matrix
4. Perform factor analysis on the data matrix
5. Interpret the factors with eigenvalues greater than 1

#### **3.4.1. Initial Keyword Selection**

To return useful information, the tool must be supplied relevant keywords (Hicks et al., 2006). Keywords were selected by examining the Eclipse Foundation's website (Eclipse.org) and the websites of five organizations selected randomly from the list of members of the Eclipse Foundation.

Keywords thought to represent the interactions of a company with the Eclipse code were selected as the initial keywords.

#### **3.4.2. Use of the Keyword Searcher Tool to Collect Data**

A list of the Uniform Resource Locators for the websites of the members of the Eclipse Foundation and the list of keywords to search for were entered into the search tool. The tool returns an occurrence matrix that contains the number of hits for each keyword on each of the sites. A hit is defined as a page that contains the keyword that is being searched for.

The keyword searcher tool returns data that is normalized based on website size (Hicks et al., 2006). This ensures that the frequency measure is independent of website size. For example, The website for Brocade (www.brocade.com) has 2,150 pages. The keywords creating OR producing appear on 289 of those pages. The normalized data used is:

$$289/2150 * 100 = 13.44\%$$

### 3.4.3. Refine Raw Data Matrix

The raw data matrix is manually examined and refined. Ubiquitous keywords (those found on all or almost all of the sample websites) and keywords with so few occurrences they do not help define a pattern are removed.

Figure 6 illustrates the method used to refine the data matrix

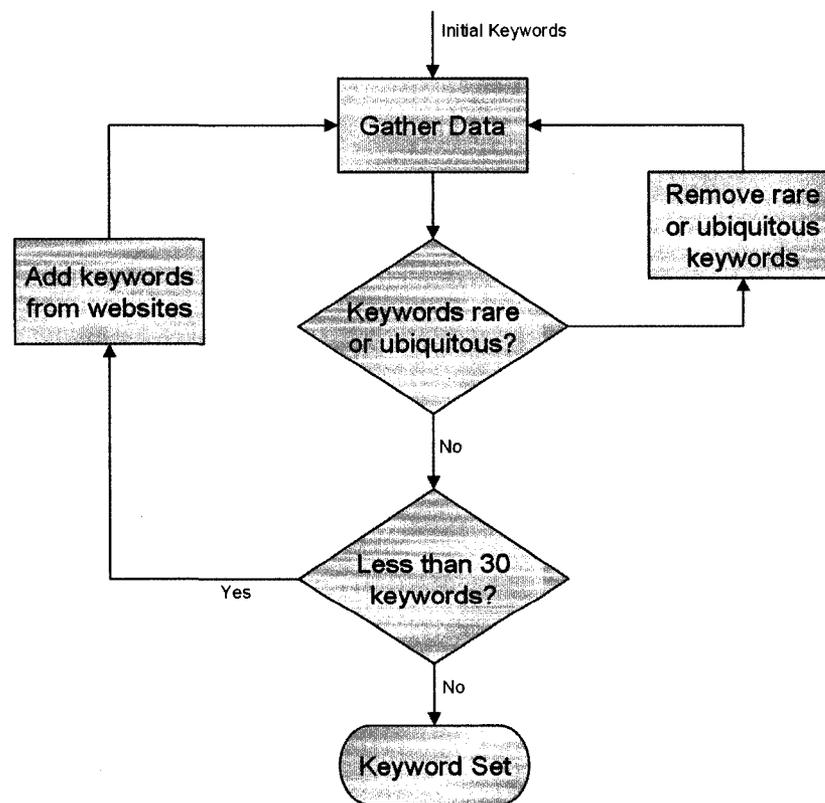


Figure 6: Method for refining data matrix

#### **3.4.4. Factor Analysis**

Factor analysis was performed on the data. Factor analysis is a complex method for analyzing data due to the large number of choices in technique to use and the few concrete guidelines. The method carried out in this research uses the best practices defined in Costello and Osborne (2005) and Darlington (2007).

SPSS was used to analyze the data. A classical factor analysis rather than a principal component analysis (Costello & Osborne, 2005) approach was taken. This is because principal component analysis is a reduction method made popular when computational power was expensive. Today's computers can easily perform proper factor analysis without any loss of data, excessive cost or long delay.

In the event that the data collected is normally distributed, the maximum likelihood method will be used to extract factors. If the data is skewed substantially, alpha extraction will be applied to the data. Alpha extraction considers analyzed variables to be a sample from the "universe of potential variables" (Hicks et al., 2006). Varimax rotation was employed in order to produce an orthogonal set of interpretable dimensions and uncorrelated factors. All factors with eigenvalues greater than 1.0 were retained for rotation.

Keywords with little or no variance were dropped. Keywords that had a low Kaiser-Meyer-Olkin Measure of Sampling Adequacy were removed from the analysis.

## **4. RESULTS**

The following section contains the results of the data analysis performed using the SPSS software package.

### **4.1. Sample**

The sample consisted of 163 Eclipse member organizations. The list of sample organizations along with their website location and type of Eclipse membership is in Appendix A: List of the 163 Eclipse Member Organizations in the Sample.

### **4.2. Keywords**

Keywords were developed using the method described in Hicks et al. (2006) using the websites for the Eclipse Foundation (2008 & 2008c), Novell, Ilog, TIBCO, Kestral and Lombardi Software. A total of 43 keywords were identified.

### **Descriptive Statistics**

The extraction method used was Alpha Extraction because the data violates the conditions of normality. Table 2 provides the descriptive statistics for the number of pages on an organization's website.

**Table 2: Descriptive statistics for the number of pages on an organization's website**

	Minimum	Maximum	Mean	Std. Deviation	Skewness	Kurtosis
Total pages on the organization's website	15	2,340,000	56,912.80	259,634.30	6.92	51.86

Table 2 shows the wide range and variation of web site sizes among the sample. This range and variation necessitated the data normalization implemented in the keyword search tool.

#### **4.4. Factor Analysis**

Nine variables were dropped because they were infrequently used and were having a negative affect on the value of the Kaiser-Meyer-Olkin Measure of Sampling Adequacy. The variables dropped were: associate member, open technology, Eclipse Public License OR epl, Mike Milinkovich, strategic member, community projects OR community project, ecosystem OR ecosystems, java OR ee OR j2ee OR j2se OR jdk OR jre OR jsf OR java server faces and "Eclipse Foundation".

Twelve variables were dropped because they did not effectively distinguish Eclipse interactions. The value of the Kaiser-Meyer-Olkin Measure of Sampling Adequacy increased after they were removed. These variables included the following keywords: collaboration OR collaborate, community, "integration platform", ide OR "development environment", "seamless integration", commercially OR commercial, "open source",

"tool platform", embracing OR embrace OR embraces, perspective, Eclipse and standardized OR standards-based OR "standards based" OR "open standards" OR standardization.

Out of the original 43 variables, 22 were used in the factor analysis. The subject to item ratio (in this case organization to keyword) was 163/22. This is greater than the requisite minimum of 2/1 (Stanek, 1993).

Table 3 shows the communalities for the variables used in the factor analysis. Initial communality represents the percentage of variance in the variable explained by all of the variables. The extracted communalities represent the percentage of variance in the variable explained by the extracted factors.

**Table 3: Communalities for the variables used in the Factor Analysis with Alpha Factoring as the extraction method (sample size = 163)**

	<b>Variable</b>	<b>Initial</b>	<b>Extraction</b>
1	"based on" OR based-on	0.620	0.583
2	"cost reduction" OR "reduction of costs" OR "lower cost"	0.455	0.432
3	"open source solutions"	0.155	0.083
4	consistent OR consistently OR consistency	0.562	0.567
5	creating OR producing	0.592	0.420
6	customize OR customized	0.503	0.390
7	equinox OR swORdfish OR smila OR riena	0.359	0.296
8	extensibility OR extensible	0.299	0.248
9	flexibility OR flexible	0.652	0.647
10	framework OR frameworks	0.468	0.556
11	gef OR emf OR gmf OR emft OR	0.529	0.855

	gmt OR m2m		
12	incorporating OR incorporate	0.465	0.458
13	invest OR investing OR investment	0.536	0.517
14	jdt OR cdt OR wtp OR dtp OR pde OR ttp	0.518	0.395
15	performance	0.585	0.693
16	platform	0.396	0.374
17	plug-ins OR plug-ins OR plug-in or plug-in	0.652	0.703
18	quickly OR quick OR fast OR faster	0.591	0.663
19	reliable OR reliability	0.475	0.373
20	scalable OR scalability	0.446	0.343
21	strategic OR strategy	0.290	0.239
22	rcp OR ercp OR babel OR ide	0.579	0.588

Table 4 contains the results of the Kaiser-Meyer-Olkin (KMO) and Bartlett's Test. A KMO of greater than 0.60 is the minimum acceptable level and a maximum of 0.01 significance is acceptable for the Bartlett's test (Hicks et al., 2006). The values for the sample data both fall within the acceptable range.

**Table 4: Kaiser-Meyer-Olkin and Bartlett's test (sample size = 163)**

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		0.829
Bartlett's Test of Sphericity	Approx. Chi-Square	1506.714
	df	231.000
	Sig.	0.000

The following figure shows the eigenvalues of the extracted factors. Four factors had eigenvalues greater than 1. This test also correlates with that used by Costello and Osborne (2005). They used a scree plot analysis to decide the number of factors to retain (see Appendix C: Factor Analysis Techniques).

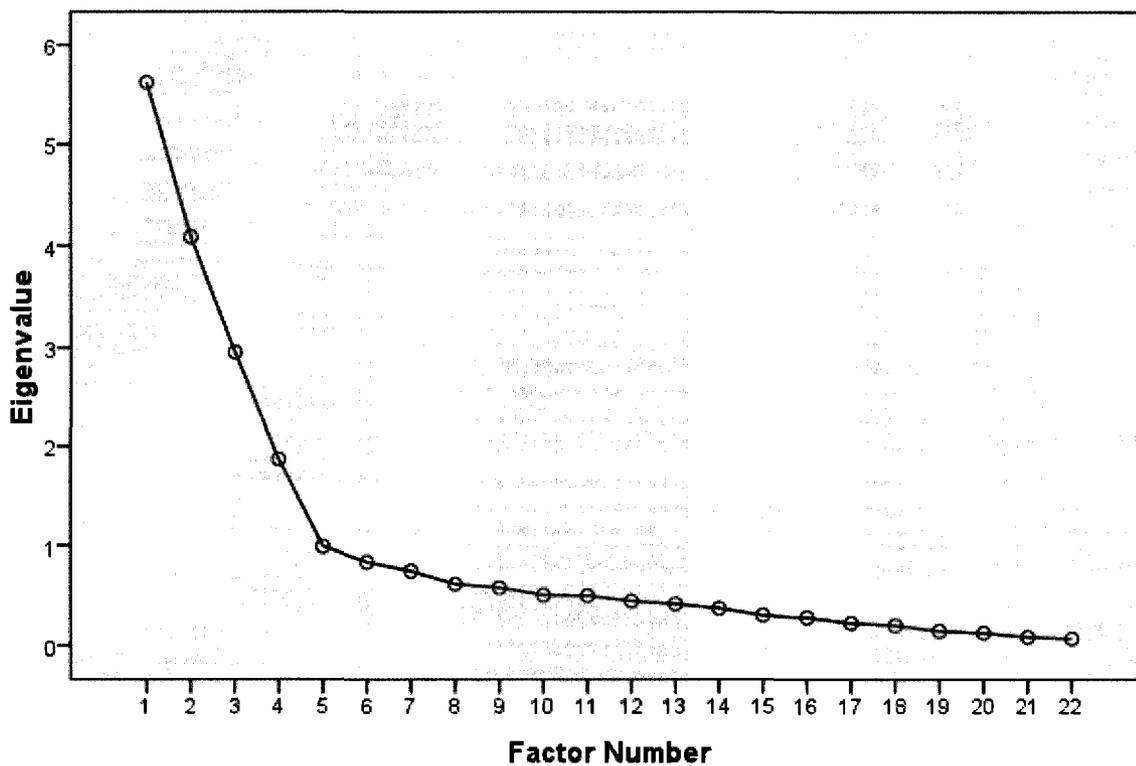


Figure 7: Scree plot for Eclipse member organizations (sample size = 163).

Table 5 shows that 59.44% of the sample's total variance was explained by the four extracted factors.

Table 5: Total Variance Explained by Extracted Factors (sample size = 163)

Factor	Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %
1	3.945	17.930	17.930
2	3.389	15.407	33.337
3	3.089	14.041	47.378
4	2.654	12.065	59.443

Table 6 shows the four factor solution. Variables with factor loadings greater than 0.4 are displayed in the table and the rest are omitted. Variables are sorted descending by factor loading.

**Table 6: Factor Analysis Results (sample size = 163)**

Interpretation	Variable		Factor Loadings			
			1	2	3	4
<b>Technology Beneficiary</b> - organization uses multiple Eclipse projects to deliver offerings	10	gef OR emf OR gmf OR emft OR gmt OR m2m	0.90			
	13	jdt OR cdt OR wtp OR dtp OR pde OR ttp	0.56			
	6	equinox OR swORdfish OR smila OR riena	0.54			
	2	customize OR customized	0.53			
	22	rcp OR ercp OR babel OR ide	0.47			
<b>Platform Beneficiary</b> - organization creates new applications based on Eclipse platform or offers services for applications	3	consistent OR consistently OR consistency		0.91		
	7	extensibility OR extensible		0.88		
	8	flexibility OR flexible		0.86		
	9	framework OR frameworks		0.70		
	16	platform		0.49		
<b>Open Source Beneficiary</b> - organization uses Eclipse as a development tool or part of a deployment stack	19	reliable OR reliability			0.69	
	18	quickly OR quick OR fast OR faster	0.52		0.66	
	4	cost reduction OR "reduction of costs" OR "lower cost"			0.62	
	15	performance		0.47	0.57	
	11	incorporating OR incorporate			0.56	
	12	invest OR investing OR investment			0.55	
	20	scalable OR scalability			0.51	
<b>Producer Beneficiary</b> - organization extends existing Eclipse applications	17	plug-ins OR plug-ins OR plug-in or plug-in	0.46			0.80
	5	creating OR producing				0.78
	1	based on OR based-on			0.45	0.68
	14	open source solutions				0.56
	21	strategic OR strategy			0.42	0.49

Extraction Method: Alpha Factoring.

Rotation Method: Varimax with Kaiser Normalization.

Rotation converged in 5 iterations.

This solution appears to be a valid solution. All four factors have more than three strongly loading variables (i.e., 0.5 or greater). Costello & Osborne (2005) state that three or more strongly loading variables in a factor are an indicator of a solid factor.

#### 4.5. Summary of the Results Applied to the 163 Eclipse Members

The websites of the 163 Eclipse member organizations were examined in order to code them based on the results of the factor analysis. Table 7 presents the organizations classified by factor. An X in one of the four factor columns means that the organization interacts with the Eclipse platform in the manner described by the factor.

**Table 7: Identification and classification of the 163 Eclipse Member Organizations**

Organization Name	Type of Eclipse Membership	Factor				Code
		1	2	3	4	
ACCESS Systems America Inc.	Solutions Member		x			2
Accurev Inc.	Solutions Member				x	4
ACM Queue	Associate Member		x			2
Actuate Corporation	Strategic Developer	x	x	x	x	1234
AdaCore Technologies	Solutions Member		x		x	24
Addison Wesley	Associate Member	x	x			12
Adobe Systems	Solutions Member		x		x	24
Advanced Micro Devices (AMD)	Solutions Member		x		x	24
Aldon	Solutions Member	x	x	x	x	1234
Algo System Co., Ltd.	Solutions Member	x	x			12
ANCiT Consulting	Solutions Member	x	x	x		123
andrena objects ag	Solutions Member	x			x	14

Anyware Technologies	Solutions Member	x	x	x	x	1234
Aptana, Inc.	Solutions Member	x			x	14
ARM Limited	Solutions Member		x		x	24
AvantSoft, Inc	Solutions Member		x			2
Band XI International	Solutions Member	x	x	x	x	1234
Black Duck Software Inc.	Solutions Member			x		3
BLU AGE	Solutions Member		x		x	24
Bluenog Corp	Solutions Member			x		3
Borland Software Corp.	Strategic Developer	x	x	x	x	1234
BREDEX GmbH	Solutions Member	x			x	14
Brocade	Solutions Member					
brox IT-Solutions GmbH	Strategic Developer	x	x	x	x	1234
BZ Media	Associate Member		x			2
CA Inc.	Strategic Consumer		x			2
CENIT AG	Solutions Member	x	x			12
Cisco Systems, Inc.	Solutions Member			x		3
Cloudsmith Inc.	Strategic Developer		x	x		23
CollabNet, Inc.	Solutions Member	x		x	x	134
Communications and Media Arts (CMA)	Associate Member		x			2
compeople AG	Solutions Member				x	4
Compuware	Solutions Member	x	x	x		123
Curl Inc.	Solutions Member		x		x	24
DDC-I	Solutions Member		x		x	24
DFKI GmbH	Associate Member		x			2
DSDM	Associate Member				x	4
EADS Deutschland GmbH	Solutions Member		x			2
Embarcadero Technologies Inc.	Solutions Member	x	x		x	124
empolis GmbH	Solutions Member		x		x	24
Enea Embedded Technology AB	Solutions Member	x	x		x	124
Ericsson AB	Solutions Member	x	x		x	124
eration A.S.	Solutions Member		x			2
Etish Limited	Solutions Member		x		x	24

ETRI (Electronics & Telecommunications Research Institute)	Associate Member				x	4
Excelsior LLC	Solutions Member		x		x	24
Exist Global (formerly DevZuz)	Strategic Developer		x	x		23
Fraunhofer Institute for Open Communication Systems (FOKUS)	Associate Member				x	4
Freescale Semiconductor	Solutions Member		x			2
froglogic GmbH	Solutions Member	x			x	14
Fujitsu Limited	Solutions Member		x			2
Genuitec, LLC	Strategic Developer	x	x	x	x	1234
Gerhardt Informatics Kft.	Solutions Member	x	x			12
Google Inc.	Solutions Member		x			2
Hewlett-Packard Company (HP)	Solutions Member			x		3
Hitachi, Ltd., Software Division	Solutions Member		x		x	24
IBM	Strategic Developer	x	x	x	x	1234
IBS AB	Solutions Member		x			2
IDG Japan	Associate Member				x	4
IFS Institut für Software	Associate Member				x	4
ILOG	Solutions Member				x	4
Ingres Corporation	Solutions Member		x	x		23
Innoopract	Strategic Developer	x	x		x	124
Innovations Softwaretechnologie GmbH	Solutions Member				x	4
Instantiations, Inc.	Solutions Member	x	x		x	124
instinctools GmbH	Solutions Member	x	x	x	x	1234
Intalio Inc.	Solutions Member		x	x		23
Intel Corporation	Strategic Consumer	x	x		x	124
Intervice Inc.	Solutions Member		x			2
IONA Technologies	Strategic Developer				x	4

itemis AG	Strategic Developer	x	x	x	x	1234
iWay Software	Solutions Member		x		x	24
Kestral Computing P/L (formerly Jiva Medical)	Solutions Member					
Klocwork	Solutions Member				x	4
KPIT Cummins Infosystems Ltd.	Solutions Member	x			x	14
Krugle, Inc.	Solutions Member			x	x	34
Laszlo Systems	Solutions Member			x	x	34
Lombardi Software	Solutions Member		x		x	24
LSI Corporation	Solutions Member				x	4
LynuxWorks Inc.	Solutions Member		x	x	x	234
Macraigor Systems LLC	Solutions Member		x		x	24
MDS Technology Co., Ltd.	Solutions Member				x	4
Mentor Graphics	Solutions Member		x		x	24
metafinanz	Solutions Member			x		3
Micro Focus	Solutions Member				x	4
MicroDoc Computersystems GmbH	Solutions Member				x	4
Mind8 GmbH	Solutions Member				x	4
Misys International Banking Systems	Solutions Member		x			2
MKS Inc.	Solutions Member	x			x	14
MontaVista	Solutions Member				x	4
Motorola	Strategic Developer	x	x		x	124
MySQL	Solutions Member			x	x	34
NEC	Solutions Member		x		x	24
NetApp	Solutions Member			x	x	34
Nexaweb Technologies	Solutions Member	x	x	x		123
nexB Inc.	Solutions Member	x	x	x	x	1234
Nokia	Strategic Developer	x	x			12
Novell	Solutions Member			x	x	34
Obeo	Strategic Developer	x	x	x	x	1234
Object Management Group	Associate Member			x		3
OC Systems	Solutions Member				x	4

Open Source Lab	Associate Member		x	x		23
Open Systems Publishing	Associate Member		x			2
OpenMake Software Inc.	Solutions Member		x	x	x	234
OpenMethods LLC	Strategic Developer		x			2
Oracle	Strategic Developer	x	x	x	x	1234
OSGi Alliance, Inc.	Associate Member	x			x	14
OW2	Associate Member	x			x	14
Palamida, Inc	Solutions Member		x	x		23
Parasoft	Solutions Member				x	4
Perforce	Solutions Member				x	4
Piterion	Solutions Member		x		x	24
Polarion Software	Solutions Member			x	x	23
Progress Software	Solutions Member		x	x		23
Prosyst Software	Solutions Member	x	x	x	x	1234
Protecode Inc.	Solutions Member			x	x	34
Purple Scout AB	Solutions Member			x	x	34
Puzzle ITC GmbH	Solutions Member		x		x	24
QNX Software Systems Co.	Solutions Member	x			x	14
Quest Software, Inc.	Solutions Member				x	4
Red Hat, Inc.	Solutions Member	x		x		13
Redmond Media Group	Associate Member					2
Remain BV	Solutions Member				x	4
Replay Solutions	Solutions Member				x	4
Research In Motion	Solutions Member			x	x	34
RTC Group Inc.	Associate Member		x			2
SalesForce.com	Solutions Member		x	x		23
SAP AG	Strategic Consumer	x		x		13
SAS	Solutions Member	x			x	14
Serena Software, Inc.	Solutions Member		x			2
Siemens	Solutions Member				x	4
SIGS Datacom	Associate Member		x		x	24
Skyway Software, Inc.	Solutions Member		x	x		23
SlickEdit Inc.	Solutions Member				x	4
Social Physics	Associate Member				x	4

Software & Support Verlag	Associate Member		x			2
Software AG	Solutions Member		x		x	24
Sony Ericsson	Solutions Member		x		x	24
Sopera GmbH	Strategic Developer	x			x	14
Soyatec	Solutions Member	x	x	x		123
SpikeSource Inc.	Solutions Member		x			2
SpringSource, Inc	Solutions Member				x	4
Standards for Technology in Automotive Retail	Associate Member	x	x			12
STMicroelectronics	Solutions Member				x	4
Sybase	Strategic Developer	x	x			12
Symbian	Solutions Member				x	4
Teamprise	Solutions Member				x	4
Tensilica	Solutions Member		x			2
Texas Instruments	Solutions Member		x			2
Thales Group	Solutions Member		x	x		23
The Medical Banking Project	Associate Member		x		x	14
The RCP Company	Solutions Member		x	x		23
TIBCO	Solutions Member		x			2
TietoEnator Telecom R&D	Solutions Member	x	x		x	124
TimeSys Corporation	Solutions Member		x			2
Toshiba Corporation	Solutions Member		x			2
VirtualLogix, Inc.	Solutions Member				x	4
Virtutech, Inc.	Solutions Member		x		x	24
Webtide	Solutions Member	x	x	x	x	1234
Weigle Wilczek GmbH	Solutions Member	x	x	x		123
Wind River	Strategic Developer	x	x	x	x	1234
Zend Technologies	Strategic Developer	x	x		x	124
Zensar Technologies, Ltd.	Solutions Member		x			2

Table 8 shows the number of organizations that share the same codes as well as a description for each code. For example, nine organizations share code 124. This means that these organizations interact with the Eclipse platform in ways described by the interpretation for factors 1, 2 and 4. This data provides information regarding the relative popularity of the different Eclipse interactions. A graphical representation of this data can be found in Figure 10.

**Table 8: Distribution of Eclipse Members by Code**

<b>Code</b>	<b>Description of the Code</b>	<b>Number of Members</b>	<b>Percent</b>
1	Organization uses multiple Eclipse projects to deliver offerings	0	0.0%
2	Organization creates new applications based on Eclipse platform	28	17.2%
3	Organization uses Eclipse as a development tool or part of a deployment stack	6	3.7%
4	Organization extends existing Eclipse applications	31	19.0%
12	Organization uses multiple Eclipse projects to deliver offerings and creates new applications based on Eclipse platform	7	4.3%
13	Organization uses multiple Eclipse projects to deliver offerings and uses Eclipse as a development tool or part of a deployment stack	2	1.2
14	Organization uses multiple Eclipse projects to deliver offerings and extends existing Eclipse applications	12	7.4%
23	Organization creates new applications based on Eclipse platform and uses Eclipse as a development tool or part of a deployment stack	12	7.4%
24	Organization creates new applications based on Eclipse platform and extends existing Eclipse applications	22	13.5%

34	Organization uses Eclipse as a development tool or part of a deployment stack and extends existing Eclipse applications	8	4.9%
123	Organization uses multiple Eclipse projects to deliver offerings, creates new applications based on Eclipse platform and uses Eclipse as a development tool or part of a deployment stack	5	3.1%
124	Organization uses multiple Eclipse projects to deliver offerings, creates new applications based on Eclipse platform and extends existing Eclipse applications	9	5.5%
134	Organization uses multiple Eclipse projects to deliver offerings, uses Eclipse as a development tool or part of a deployment stack and extends existing Eclipse applications	1	0.6%
234	Organization creates new applications based on Eclipse platform, uses Eclipse as a development tool or part of a deployment stack and extends existing Eclipse applications	2	1.2%
1234	Organization uses multiple Eclipse projects to deliver offerings, creates new applications based on Eclipse platform, uses Eclipse as a development tool or part of a deployment stack and extends existing Eclipse applications	16	9.8%

## **5. DISCUSSION OF RESULTS**

This chapter has seven sections. Section 5.1 interprets the factors that were. Section 5.2 organizes the organizations in the sample into clusters based on the way they interact with the Eclipse platform. Section 5.3 compares the way the Eclipse members interact with the platform with how they represent their interactions on their websites. Sections 5.4 and 5.5 discuss members that develop new Eclipse applications or use it as open source and members that use multiple Eclipse technologies or extend existing Eclipse applications. Section 5.6 provides a comparison between Carbone's (2007) open source maturity model and the results obtained in this research. Section 5.7 examines the distribution of the factors by Eclipse membership type.

### **5.1. Interpretation of the Factors**

The factor analysis identified four factors with eigenvalues greater than one. Table 9 shows the four factors and the researcher's interpretation of the factors' meaning and rationale. The researcher believes that the factors relate to the way Eclipse members interact with the Eclipse platform to create their offerings.

**Table 9: Interpretation of the Extracted Factors**

<b>Factor</b>	<b>Interpretation</b>	<b>Rationale</b>
1	<p><b>Technology Beneficiary</b></p> <p>Organization uses multiple Eclipse projects to deliver offerings</p>	The organization develops for and integrates a number of Eclipse technologies in order to create their offering. This integration provides a competitive advantage because of development cost savings and an increase in features.
2	<p><b>Platform Beneficiary</b></p> <p>Organization creates new applications based on Eclipse platform or offers services for applications</p>	The organization takes advantage Eclipse as a platform by releasing new applications based on the Eclipse platform or another Eclipse application, or the organization offers services needed by Eclipse platform users (such as training or documentation).
3	<p><b>Open Source Beneficiary</b></p> <p>Eclipse used as a development tool or for deployment; open source business model used</p>	The organization takes advantage of the open source nature of Eclipse by using Eclipse to save on costs or by offering free Eclipse plug-ins or Eclipse-based applications and charging for complimentary services (support, maintenance, etc.).
4	<p><b>Producer Beneficiary</b></p> <p>Organization extends existing Eclipse applications</p>	The organization develops Eclipse plug-ins to enhance Eclipse projects or other Eclipse-based applications.

## 5.2. Organization Clusters

Figure 8 and Figure 9 show the Eclipse member organizations clustered by the descriptions in Table 8. The number above each box represents the code (from Table 8) and in parenthesis is the number of organizations in the cluster. The names of the companies in the cluster are included in the box itself.

Figure 8: Eclipse Member Organizations Grouped by the Codes Described in Table 8 (sample size = 163)

2 (28)

ACCESS Systems Amdex Inc., AUM Queue, AvantiSoft, Inc. B2 Media, CA Inc., Communications and Media Arts (CMA), DFKI GmbH, EADS Deutschland GmbH, eStation A.S., Freescale Semiconductor, Fujitsu Limited, Google, Inc., IBS AD, Interwise Inc., Infos International Jenteng Systems, Open Systems Publishing, OpenMethods LLC, Reinhard Media Group, RTC Group Inc., Secura Software, Inc., Software & Support Verlag, SpikeSource Inc., Tensilica, Texas Instruments, TIBCO, Tiscali Systems Corporation, Toshiba Corporation, Zensor Technologies, Ltd.

3 (6)

Black Duck Software Inc., Bluewin Corp., Cisco Systems, Inc., Hewlett-Packard Company (HP), metafinanz, Object Management Group

4 (31)

Arcusys Inc., compeople AG, DSDM, ETRI (Electronics & Telecommunications Research Institute), Fraunhofer Institute for Open Communication Systems (FOKUS), IDG Japan, IPS Institut für Software, ILOG, Innovations Softwaretechnologie GmbH, IONA Technologies, Klocwork, LSI Corporation, MDS Technology Co., Ltd., Micro Focus, MicroDoc Computersystems GmbH, MinidB GmbH, MontaVista, OC Systems, Parallel, Perfence, Quest Software, Inc., Pomain BV, Reply Solutions, Siemens, ShickEdit Inc., Social Physics, SpringSource, Inc., STMicroelectronics, Syntian, Teamonix, VirtualLogic, Inc.

12 (7)

Action Wesley, Algo System Co., Ltd., CENIT AG, Uehardt Informatik EA, Nokia, Standards for Technology in Automotive Retail, Sybase

13 (2)

Red Hat, Inc., SAP AG

14 (12)

andrea objects ag, Argona, Inc., BREDEX GmbH, froglogic GmbH, KHT Curriers Infosystems Ltd., MKS Inc., OSO Alliance, Inc., OW2, QNX Software Systems Co., SAS, Sopera GmbH, The Medical Banking Project

23 (12)

Cloudsmith Inc., Enist Global (formerly DevZup), Ingres Corporation, Intalio Inc., Open Source Lab, Palantir, Inc., Polaris Software, Progress Software, Salesforce.com, Skyway Software, Inc., Thales Group, The RCP Company

Figure 9: Eclipse Member Organizations Grouped by the Codes Described in Table 8 (sample size = 163) (continued)

24 (22)

AdaCore Technologies, Adobe Systems, Advanced Micro Devices (AMD), ARM Limited, BLU AGE, Cavi Inc., DDC-I, empolis GmbH, Elish Limited, Excelsior LLC, Hitachi Ltd, Software Division, iWay Software, Lombardi Software, Macrigger Systems LLC, Mentor Graphics, NEC, Piterion, Purzle ITC GmbH, SIOS Datacom, Software AG, Sony Ericsson, Virtutech, Inc.

34 (8)

Krugle, Inc., Laszlo Systems, MySQL, NetApp, Novell, Protecode Inc., Purple Scout AB, Research In Motion

123 (5)

ANCIT Consulting, Compuware, Nexaweb Technologies, Soyatec, Weigle Wilczek GmbH

124 (9)

Embarcadero Technologies Inc., Enes Embedded Technology AB, Ericsson AB, Innocrypt, Instantions, Inc, Intel Corporation, Motorola, Neteon or Telecom R&D, Zend Technologies

134 (1)

CollabNet, Inc.

234 (2)

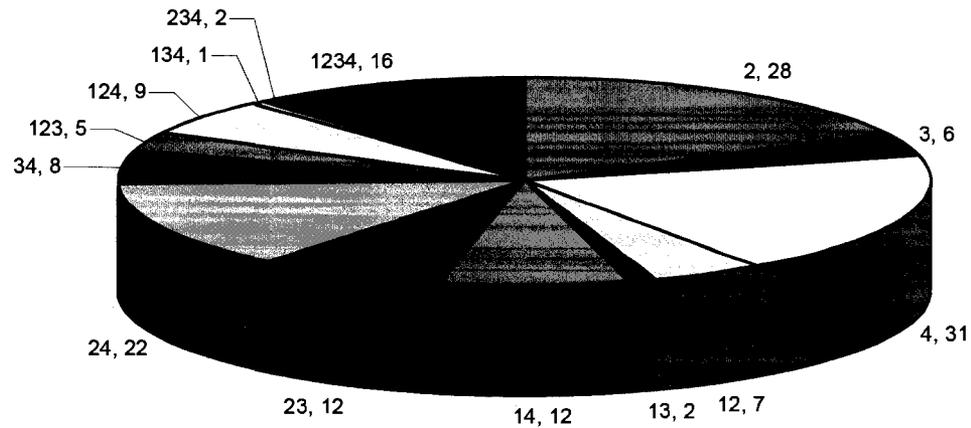
LynuxWorks Inc., OpenMake Software Inc.

1234 (16)

Actuate Corporation, Aldon, Anyware Technologies, Band XI International, Bodand Software Corp., brox IT Solutions GmbH, Geraated, LLC, IBM, instinctools GmbH, items AG, nexB inc., Obco, Oracle, Paeryst Software, Webtide, Wind River

**Figure 10: Distribution of Organizations Grouped By Codes Described in Table**

**8**



Fifty two of the 163 members (34%) interact with multiple Eclipse projects. These organizations are denoted by clusters 12, 13, 14, 123, 124, 134 and 1234. None of the organizations rely solely on the technological aspect of Eclipse in their interactions, rather they use multiple technologies to create new applications, extend existing applications or offer services based on multiple Eclipse projects.

One hundred and one of the 163 members (62%) interact with Eclipse by creating new Eclipse-based applications. These members are denoted by clusters 2, 12, 23, 24, 123, 124, 234 and 1234. Of the 101 members, 35 of them (21% of the total sample) interact with Eclipse by creating new Eclipse-based applications and using Eclipse as a development tool or part of a deployment stack.

Fifty two members (34% of the sample) interact with Eclipse's open source ecosystem. These organizations are denoted by clusters 3, 13, 23, 34, 123, 134, 234 and 1234. More than half of these members (27 or 17% of the total sample) interact with the open source aspect of Eclipse and develop plug-ins for existing applications.

One hundred and one of the 163 members (62%) interact with the Eclipse platform by developing plug-ins for existing applications. These organizations are denoted by clusters 4, 14, 24, 34, 123, 134, 234 and 1234. Thirty one of these members (19% of the total sample) interact with Eclipse only by developing plug-ins for existing applications.

### **5.3. Comparing Results with Member Websites**

Using the raw data returned from the keyword search tool we can examine if members accurately describe their interactions with Eclipse on their websites.

Table 10 presents the factor loadings for each of the 14 clusters that the sample Eclipse members belong to. For each cluster, the percentage of pages with at least one of the keywords for its respective factor is shown. At the bottom of the table, the average percentage of pages with at least one of the factor's keywords is shown for each factor.

**Table 10: Factor Loading by Clusters – A Comparison of Website data to Organization’s Eclipse Interactions**

Cluster	Factor			
	1	2	3	4
2		22%		
3			12%	
4				4%
12	4%	12%		
13	3%		5%	
14	5%			7%
23		37%	21%	
24		24%		4%
34			27%	4%
123	12%	25%	16%	
124	6%	22%		6%
134	3%		5%	2%
234		24%	20%	3%
1234	15%	29%	17%	6%
<b>Average</b>	7%	24%	15%	5%

This data shows how well Eclipse members represent their interactions with Eclipse on their websites. Factors 2 and 3 (creating new Eclipse applications or interacting with the Eclipse open source ecosystem) seem very well represented across all of the clusters. The exception to this is that factor three is not particularly well represented (5%) by companies that use multiple Eclipse technologies and interact with the Eclipse open source ecosystem (cluster 13 and 134).

On average, factor 1 is poorly represented relative to factors 2 and 3. This means that Eclipse members who interact with Eclipse by using use multiple Eclipse tend no to

represent this fact as prominently on their corporate websites. The exception to this trend is for clusters 123 and 1234. The data shows that organizations that use multiple Eclipse technologies, develop new Eclipse applications and use Eclipse as open source represent all three of those factors well on their corporate websites.

Eclipse members in the sample that interact with Eclipse by extending existing applications tend not to represent this information on their corporate websites as prominently. This seems to be the case across the entire sample with a maximum of 7% of pages (for cluster 14) on a member's website having words relating to this factor.

Finally, Eclipse members in the sample that create new Eclipse applications interact with Eclipse as open source (cluster 23) tend to represent their Eclipse interactions on their websites better than any other cluster.

#### **5.4. Developing New Applications and Interacting With Eclipse as Open Source**

One hundred and eighteen of the sample Eclipse members (72%) extend existing Eclipse applications or use Eclipse as open source. Members in the sample also represented these interactions very well on their websites. Iansiti & Levien (2004) state that ecosystem members must create more value than they take if their ecosystem is to grow and thrive. By giving cues as to how they benefit from Eclipse, these members help strengthen the Eclipse ecosystem.

### **5.5. Interaction with Multiple Projects or Extension of Existing Applications**

Of the 163 Eclipse members in the sample, 115 of them (71%) interact with multiple Eclipse projects or extend existing Eclipse-based applications. According to the results presented, these factors are not as well represented on corporate websites of the Eclipse members employing them. There could be a number of reasons why this is the case however the data cannot give a definitive reason. Even though these factors are not as well represented as those in section 5.4, all of the factors are still represented on at least one out of every 20 pages in an organization's website on average.

### **5.6. Comparison with Carbone's (2007) Open Source Interactions**

The four identified interactions between Eclipse members and the Eclipse platform can be compared to Carbone's five open source interactions. Table 11 provides the comparison.

**Table 11: Comparison of Eclipse Interactions with Carbone's (2007) Open****Source Interactions****Levels of Open Source Maturity**

<b>Use</b>	<b>Contribute</b>	<b>Champion</b>	<b>Collaborate</b>	<b>Redefine</b>
Use and promotion of open source software	Contributes resources to development of open source project	Contributes leadership, mobilizes developers, coordinates development	Collaboration across multiple projects to change market in organization's favour	Change value proposition of offer delivered to customers

**Eclipse Interactions**

<b>Open Source</b>	<b>Producer</b>	<b>Platform</b>	<b>Technology</b>
Eclipse used as a development tool or for deployment; open source business model used	Organization extends existing Eclipse applications	Organization uses Eclipse as a platform to build new applications or offer services	Organization uses multiple Eclipse projects to deliver offerings

The comparison shows that similarities are found between Carbone's use interaction and the open source interaction because in both cases the organization makes use of open source.

The comparison draws a connection between Carbone's contribute interaction and the producer interaction. In this case, both organizations are contributing resources to the development of an existing project or application.

The comparison finds similarities between Carbone's champion interaction and the platform interaction. In both of these cases, the organization leads development of an application based on open source and drives adoption of that application.

Finally, the comparison finds similarities between Carbone's collaborate interaction and the technology interaction. The similarity is that both organizations interact with multiple projects in order to deliver their offerings.

### **5.7. Distribution of Clusters by Eclipse Membership Type**

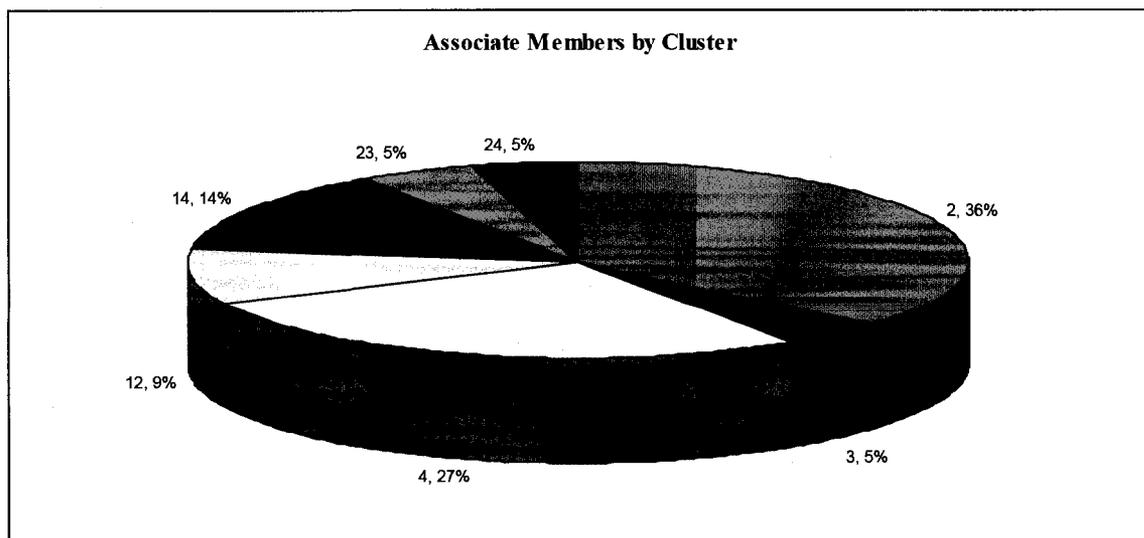
There are four types of Eclipse membership open to organizations. They are: associate member, solutions member, strategic consumer, and strategic developer (a full explanation of these membership types can be found in section 2.2.2.1).

Fifteen interaction clusters were identified in Table 8. These clusters represent a single factor, or a combination of 2, 3 or 4 factors. Table 12 provides a distribution of the sample Eclipse members distributed by their Eclipse membership type and the interaction cluster they belong to. Each data point represents the percentage (and number in parenthesis) of sample members in each of the membership types that are in a particular interaction cluster. For example 36% of associate members in the sample are platform beneficiaries.

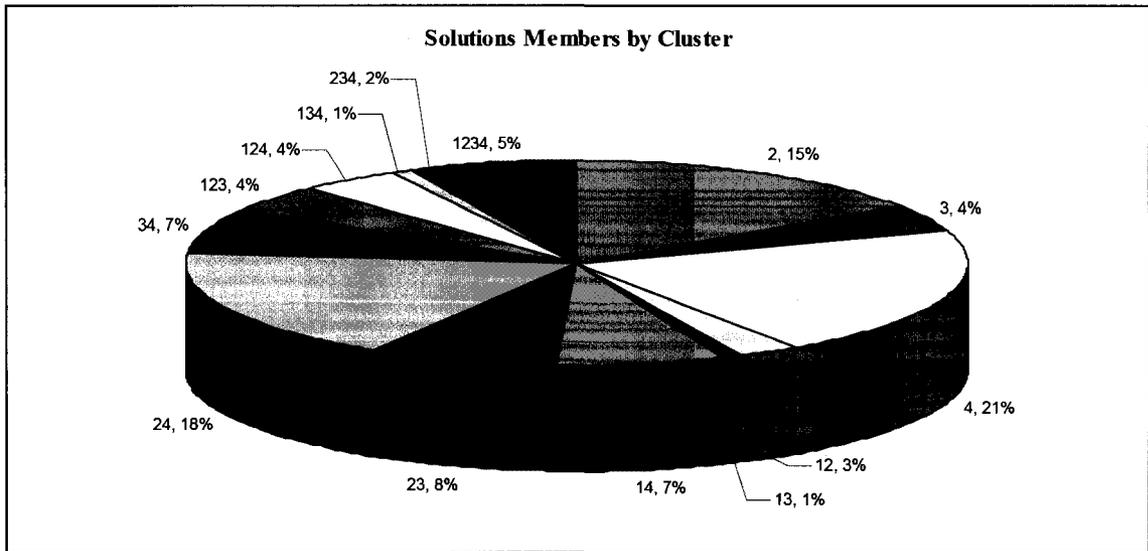
**Table 12: Distribution of Interaction Clusters by Eclipse Membership Type**

Code	Type of Eclipse Membership				
	Associate Member (22)	Solutions Member (119)	Strategic Member (Consumers & Producers) (22)	Strategic Consumer (3)	Strategic Developer (19)
1					
2	36% (8)	15% (18)	9% (2)	33% (1)	5% (1)
3	5% (1)	4% (5)			
4	27% (6)	21% (24)	5% (1)		5% (1)
12	9% (2)	3% (3)	9% (2)		11% (2)
13		1% (1)	5% (1)	33% (1)	
14	14% (3)	7% (8)	5% (1)		5% (1)
23	5% (1)	8% (9)	9% (2)		11% (2)
24	5% (1)	18% (21)			
34		7% (8)			
123		4% (5)			
124		4% (5)	18% (4)	33% (1)	16% (3)
134		1% (1)			
234		2% (2)			
1234		5% (6)	41% (9)		47% (9)

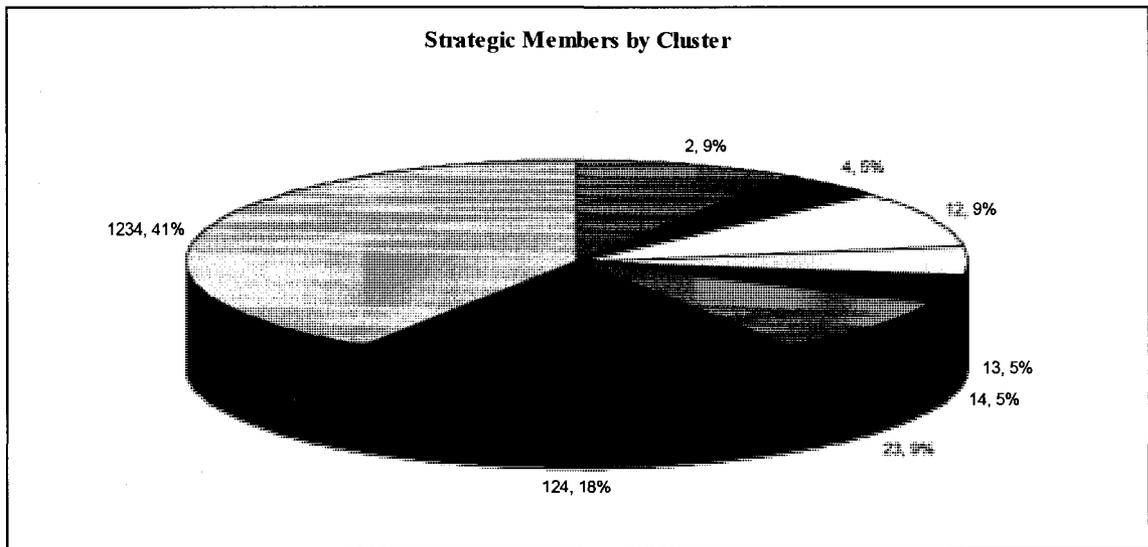
Figures Figure 11: Distribution of Associate Members by Interaction Clusters, Figure 12 and Figure 13 provide graphical representations of the data found in Table 12.



**Figure 11: Distribution of Associate Members by Interaction Clusters**



**Figure 12: Distribution of Solutions Members by Interaction Clusters**



**Figure 13: Distribution of Strategic Members by Interaction Clusters**

## **6. CONCLUSIONS**

This chapter has three sections. Section 6.1 presents the conclusions for this research. Section 6.2 identifies the limitations of this research and finally Section 6.3 provides suggestions for future research.

### **6.1. Conclusions**

This research uses a data mining technique in order to classify Eclipse members by the way they represent their relationship with Eclipse on their websites.

The main conclusions that can be drawn from the results are:

1. Eclipse member's websites provide a bounty of information that can be used by the Eclipse foundation to create value propositions.
2. Eclipse members interact with the Eclipse platform in four ways: (i) harnessing multiple Eclipse projects in order to create a competitive advantage in the market; (ii) using Eclipse as a platform to build robust, extensible applications or to provide complimentary services such as training or documentation; (iii) taking advantage of Eclipse's vast open source ecosystem using one of the many open source business models; and (iv) developing plug-ins for use in existing Eclipse applications.
3. The largest groups of Eclipse members belong in clusters 2, 4, and 24. These clusters represent platform beneficiaries, producer beneficiaries and organizations using both platform and producer interactions.

## **6.2. Limitations**

There are at least three limitations to this research. Three identified limitations are:

- The classification was only validated by two individuals. While both are very knowledgeable about Eclipse, a more systematic approach and input from company managers are required to validate the classification
- Secondary data sources are used which are not necessarily accurate indicators of an organization's values and motives
- There is a strong dependency on the selected keywords; different keywords could yield different results
- The results of a factor analysis rely on the interpretation of the extracted factors made by the researcher. This interpretation could be biased or incorrect.

## **6.3. Future Research Opportunities**

Opportunities for future research include:

- Investigation into automation of earlier data gathering procedures. Can keyword selection be automated?
- Test sensitivity of keyword selection. Do keywords selected by different researchers yield similar results?
- Investigate correlation between the types of offerings an organization produces and the interactions they use in the Eclipse ecosystem.
- Investigate financial benefits of the four interpreted interactions relative to each other (market share, market growth, return on investment, etc.)
- Conduct a human validation of results using industry experts or interviews with Eclipse member representatives to ensure accurateness of results.

- Correlation of interpreted factors. How likely is it an organization will use interaction  $x$  if they use interaction  $y$ ?
- Investigate the robustness and reliability of the Keyword Searcher Tool

## REFERENCES

- Amazon. 2008. **Amazon Simple Storage Service**. Last Accessed: October 9, 2008. Available at: <http://aws.amazon.com/s3/>.
- Basole, R. & B. Rouse. 2008. Complexity of service value networks: Conceptualization and empirical investigation. **IBM Systems Journal: Service Science, Management, and Engineering**, 47(1):53-70.
- Bauer, A., & M. Pizka. 2003. **The contribution of free software to software evolution**. Proceedings of 6th International Workshop on Principles of Software Evolution.
- Bauer, C. & A.Scharl. 2000. Quantitive evaluation of Web site content and structure. **Internet Research: Electronic Networking Applications and Policy**, 10(1):31-44.
- Bessen, J. 2005. **Open Source Software: Free Provision of Complex Public Goods**. Last accessed: October 10, 2008. Available at: <http://www.researchoninnovation.org/>.
- Carbone, P. 2006. **Competing With Open Source: Software Insights from Recent Research**. OCRI Conference "Competing with Open Source Software". Available at: <http://www.ocri.ca/events/presentations/partnership/April2106/Carbone.pdf>.
- Carbone, P. 2007. **Value Derived from Open Source is a Function of Maturity Levels**. OCRI Conference "Alchemy of Open Source Businesses", April 19, 2007.
- Christensen, C. & S. Anthony. 2004. *Seeing What's Next: Using the Theories of Innovation to Predict Industry Change*. EA Roth Publishing. P. 63.
- Christensen, C. & R. Rosenbloom. 1995. Explaining the attacker's advantage: Technological paradigms, organizational dynamics, and the value network. **Research Policy**, 24(2):233-257.
- Christensen, C. 2003. *The Innovator's Dilemma: The Revolutionary Book that Will Change the Way You Do Business*. Collins Publishers. P. 53-56.
- Costello, A., & J. Osborne. 2005. Best Practices in Exploratory Factor Analysis: Four Recommendations for Getting the Most From Your Analysis. **Practical Assessment Research & Evaluation**, 10(7):1-9.
- Darlington, R. 2007. **Factor Analysis**. Last Accessed: March 28, 2007. Available at: <http://www.psych.cornell.edu/Darlington/factor.htm>

Dholakia, R.; Miao, Z.; Dholakia, N. & D. Fortin. 2000. **Interactivity and Revisits to Websites: A Theoretical Framework**. Last Accessed: October 8, 2008. Available at: <http://ritim.cba.uri.edu/wp/>.

E\*Trade Financial. 2008. **E\*Trade Financial**. Last Accessed: October 9, 2008. Available at: <https://us.etrade.com/e/t/home>.

Eclipse Foundation. 2008. **About the Eclipse Foundation**. Last accessed: September 22, 2008. Available at: <http://www.eclipse.org/org/>.

Eclipse Foundation. 2008b. **Eclipse Public License Version 1.0**. Last Accessed: October 8, 2008. Available at: <http://www.eclipse.org/legal/epl-v10.html>.

Eclipse Foundation. 2008c. **Types of Membership**. Last Accessed: October 8, 2008. Available at: [http://www.eclipse.org/membership/become\\_a\\_member/membershipTypes.php](http://www.eclipse.org/membership/become_a_member/membershipTypes.php).

Eclipse Foundation. 2008d. **Eclipse Platform Overview**. Last accessed: November 16, 2008. Available at: <http://www.eclipse.org/platform/overview.php>.

Fabrigar, L.; Wegener, D.; MacCallum, R.; & E. Strahan. 1999. Evaluating the use of exploratory factor analysis in psychological research. **Psychological Methods**, 4(3):272-299.

Fitzgerald, B. 2006. Transformation of Open Source Software. **MIS Quarterly**, 30(3):587-598

Google. 2008. **Advertise your business on Google**. Last Accessed: October 9, 2008. Available at: <http://adwords.google.com/>

Google. 2008b. **Google AJAX Search API**. Last accessed: October 10, 2008. Available at: <http://code.google.com/apis/ajaxsearch/>.

Gustavsen, P. & E. Tilley. 2003. **Public relations communication through corporate websites: Towards an understanding of the role of interactivity**. Last Accessed: August 16, 2008. Available at: [http://praxis.massey.as.nz/issue\\_1.html](http://praxis.massey.as.nz/issue_1.html).

Hicks, D.; Libaers, D.; Porter, L. & D. Schoeneck. 2006. Identification of the Technology Commercialization Strategies of High-tech Small Firms. **Small Business Research Summary**. December. Available at: <http://www.sba.gov/advo/research/rs289tot.pdf>.

Iansiti, M. & Levien, R. 2004. *The Keystone Advantage: What the New Dynamics of Business Ecosystems Mean for Strategy, Innovation, and Sustainability*. Harvard Business School Publishing.

Insch, A. 2008. Online communication of Corporate Environmental Citizenship: A study of New Zealand's electricity and gas retailers. **Journal of Marketing Communications**, 14(2):139-153.

Koenig, J. 2004. Seven open source business strategies for competitive advantage. **IT Manager's Journal**, May 14.

Kosla, R. & H. Blockeel. 2000. **Web mining research a survey**. Proceedings of Special Interest Group on Knowledge Discovery and Data Mining, Volume 2.

Krippendorff, K. 2004. *Content Analysis: An Introduction to Its Methodology*. SAGE Publications. P. 18.

Learner, J. & J. Tirole. 2002. Some Simple Economics of Open Source. **The Journal of Industrial Economics**, L(2):192-235.

Lewin, R. & B. Regine. 1999. On the Edge in the World of Business. In Lewin, R. *Complexity: Life at the Edge of Chaos*. The University of Chicago Press. P. 197-211.

Lussier, S. 2004. New Tricks: How Open Source Changed the Way My Team Works. **Software, IEEE**, 21(1):68-72.

McGregor, J. 2008. An Increase in Value. **Journal of Object Technology**, 7(3):7-16. Available at: [http://www.jot.fm/issues/issue\\_2008\\_03/column1/](http://www.jot.fm/issues/issue_2008_03/column1/)

McMillan, S. 2000. The microscope and the moving target: The challenge of applying content analysis to the world wide web, **Journalism and Mass Communication Quarterly**, 77(1):80-98.

Moore, J. 1996. *The Death of Competition: Leadership & Strategy in the Age of Business Ecosystems*. Harper Business Publishing. P. 9, 25-26

Moore, J. 1998. The Rise of a New Corporate Form. **Washington Quarterly**. 21(1):167-181.

Merilinna, J. & M. Martinlassi. 2006. **State of the art and practice of open source component integration**. Proceedings of 32<sup>nd</sup> Euromicro Conference on Software Engineering and Advanced Applications.

- Novell. 2008. **SUSE Linux Enterprise**. Last accessed October 23, 2008. Available at: <http://www.novell.com/linux/>.
- Opoku, R. 2005, Communication of Brand Personality by Some Top Business Schools Online, Licentiate Thesis, Lulea University of Technology, ISSN: 1402-1757.
- OSGI. 2008. **OSGi Alliance Specifications**. Last accessed: October 8, 2008. Available at: <http://www.osgi.org/Specifications/HomePage>.
- Peltoniemi, M. & E. Vuori. 2004. **Business ecosystem as the new approach to complex adaptive business environments**. Proceedings of Frontiers of E-Business Research 2004.
- Perry, M. & C. Bodkin, 2000, Content analysis of 100 company Web sites. **Corporate Communications: An International Journal**, 5(2):87–97.
- Polancic, G.; Horvat, R. & T. Rozman. 2004. **Comparative Assessment of Open Source Software Using Easy Accessible Data.**, Proceedings of Information Technology Interfaces.
- Pollach, I. 2005. Corporate self-presentation on the WWW: Strategies for enhancing usability, credibility and utility. **Corporate Communications: An International Journal**, 10(4):285-301.
- Porter, A.; Schoeneck, D.; Frey, P.; Hicks, D. & D. Libaers. 2007. Mining the Internet for Competitive Technical Intelligence. **Competitive Intelligence Magazine**, 10(5):24-28
- Power, T. & G. Jerjian. 2001. *Ecosystem: Living the 12 Principles of Networked Business*. Financial Times/Prentice Hall Press.
- Raymond, E. 1999. **The Cathedral and the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary**. O'Reilly Press.
- Rennie, J. & A. McCallum, 1999. **Using reinforcement learning to spider the web efficiently**. Proceedings of the 16th International Conference on Machine Learning.
- Scacchi, W.; Feller, J.; Fitzgerald, B.; Hissam, S. & K. Lakhani. 2006. Understanding Free/Open Source Software Development Processes. **Software Process: Improvement and Practice**, 11(1):95-105
- Sharp, L. 2002. Positive response action: The ultimate goal of website communication. **Journal of Communication Management**, 6(1):41-52.

Smith, D. & M. Milinkovich. 2007. Eclipse: A Premier Open Source Community.

**Open Source Business Resource**, July, 2007. Available at:

<http://www.osbr.ca/ojs/index.php/osbr/article/viewArticle/352/313>

Stamatakis, K.; Karkaletsis, V.; Paliouras, G.; J. Horlock; Griver, C.; Curran, J. & S.

Dingare. 2003. **Domain-specific Web site identification: the CROSSMARC focused Web crawler**. Proceedings of the 2nd International Workshop on Web Document Analysis.

Stanek, D. 1993. Modeling perceptions and preference of home-based and center-based telecommuting. M.Sc Thesis, Department of Civil and Environmental Engineering, University of California, Davis, USA. Available at:

<http://www.its.ucdavis.edu/telecom/r11/factan.html>.

Sullivan, J. 1999. What are the Functions of Corporate Home Pages? **Journal of World Business**, 34(2):193-210.

Van Riel, C. & J. Balmer. 1997. Corporate identity: the concept, its measurement and management. **European Journal of Marketing**, 31(5):340-355

Wang, H. & C. Wang. 2001. Open source software adoption: a status report. **Software, IEEE**, 18(2):90-95

Weare, C. & W. Lin. 2000. Content analysis of the World Wide Web-Opportunities and challenges. **Social Science Computer Review**, 18(3):272-292.

West, J. & S. Gallagher. 2006. "Patterns of Open Innovation in Open Source Software" in *Open Innovation: Researching a New Paradigm*. Oxford University Press.

Zachary, R. 2006. Six Options For Open-Source Support. **Information Week**, March 1. Available at:

<http://www.informationweek.com/news/showArticle.jhtml?articleID=180207025>

## 7. APPENDICES

### 7.1. Appendix A: List of the 163 Eclipse Member Organizations in the Sample

	<b>Organization Name</b>	<b>Website Address</b>	<b>Type of Eclipse Membership</b>
1	ACCESS Systems America Inc.	<a href="http://www.access-company.com/">http://www.access-company.com/</a>	Solutions Member
2	Accurev Inc.	<a href="http://www.accurev.com/">http://www.accurev.com/</a>	Solutions Member
3	ACM Queue	<a href="http://www.acmqueue.com/">http://www.acmqueue.com/</a>	Associate Member
4	Actuate Corporation	<a href="http://www.birt-exchange.com/">http://www.birt-exchange.com/</a>	Strategic Developer
5	AdaCore Technologies	<a href="http://www.adacore.com/">http://www.adacore.com/</a>	Solutions Member
6	Addison Wesley	<a href="http://awprofessional.com/">http://awprofessional.com/</a>	Associate Member
7	Adobe Systems	<a href="http://www.adobe.com/">http://www.adobe.com/</a>	Solutions Member
8	Advanced Micro Devices (AMD)	<a href="http://developer.amd.com/">http://developer.amd.com/</a>	Solutions Member
9	Aldon	<a href="http://www.aldon.com/">http://www.aldon.com/</a>	Solutions Member
10	Algo System Co., Ltd.	<a href="http://www.algosystem.co.jp/">http://www.algosystem.co.jp/</a>	Solutions Member
11	ANCiT Consulting	<a href="http://www.ancitconsulting.com/">http://www.ancitconsulting.com/</a>	Solutions Member
12	andrena objects ag	<a href="http://www.andrena.de/">http://www.andrena.de/</a>	Solutions Member
13	Anyware Technologies	<a href="http://www.anyware-tech.com/">http://www.anyware-tech.com/</a>	Solutions Member
14	Aptana, Inc.	<a href="http://www.aptana.com/">http://www.aptana.com/</a>	Solutions Member
15	ARM Limited	<a href="http://www.arm.com/">http://www.arm.com/</a>	Solutions Member
16	AvantSoft, Inc	<a href="http://www.avantsoft.com/">http://www.avantsoft.com/</a>	Solutions Member
17	Band XI International	<a href="http://www.bandxi.com/">http://www.bandxi.com/</a>	Solutions Member
18	Black Duck Software Inc.	<a href="http://www.blackducksoftware.com/">http://www.blackducksoftware.com/</a>	Solutions Member
19	BLU AGE	<a href="http://www.bluage.com/">http://www.bluage.com/</a>	Solutions Member
20	Bluenog Corp	<a href="http://www.bluenog.com/">http://www.bluenog.com/</a>	Solutions Member
21	Borland Software Corp.	<a href="http://www.borland.com/">http://www.borland.com/</a>	Strategic Developer
22	BREDEX GmbH	<a href="http://www.bredex.de/en/">http://www.bredex.de/en/</a>	Solutions Member
23	Brocade	<a href="http://www.brocade.com/">http://www.brocade.com/</a>	Solutions Member
24	brox IT-Solutions GmbH	<a href="http://www.brox.de/en/">http://www.brox.de/en/</a>	Strategic Developer
25	BZ Media	<a href="http://www.bzmedia.com/">http://www.bzmedia.com/</a>	Associate Member
26	CA Inc.	<a href="http://www.ca.com/">http://www.ca.com/</a>	Strategic Consumer
27	CENIT AG	<a href="http://www.cenit.de/">http://www.cenit.de/</a>	Solutions Member
28	Cisco Systems, Inc.	<a href="http://www.cisco.com/">http://www.cisco.com/</a>	Solutions Member
29	Cloudsmith Inc.	<a href="http://www.cloudsmith.com/">http://www.cloudsmith.com/</a>	Strategic Developer
30	CollabNet, Inc.	<a href="http://www.collab.net/">http://www.collab.net/</a>	Solutions Member
31	Communications and Media Arts (CMA)	<a href="http://cma.dyndns.tv/">http://cma.dyndns.tv/</a>	Associate Member
32	compeople AG	<a href="http://www.compeople.de/">http://www.compeople.de/</a>	Solutions Member

33	Compuware	<a href="http://www.compuware.com/">http://www.compuware.com/</a>	Solutions Member
34	Curl Inc.	<a href="http://www.curl.com/">http://www.curl.com/</a>	Solutions Member
35	DDC-I	<a href="http://www.ddci.com/">http://www.ddci.com/</a>	Solutions Member
36	DFKI GmbH	<a href="http://www.dfki.de/web/">http://www.dfki.de/web/</a>	Associate Member
37	DSDM	<a href="http://www.dsdm.org/">http://www.dsdm.org/</a>	Associate Member
38	EADS Deutschland GmbH	<a href="http://www.eads.com/">http://www.eads.com/</a>	Solutions Member
39	Embarcadero Technologies Inc.	<a href="http://www.embarcadero.com/">http://www.embarcadero.com/</a>	Solutions Member
40	empolis GmbH	<a href="http://www.empolis.com/">http://www.empolis.com/</a>	Solutions Member
41	Enea Embedded Technology AB	<a href="http://www.enea.com/">http://www.enea.com/</a>	Solutions Member
42	Ericsson AB	<a href="http://www.ericsson.com/">http://www.ericsson.com/</a>	Solutions Member
43	eteration A.S.	<a href="http://www.eteration.com/">http://www.eteration.com/</a>	Solutions Member
44	Etish Limited	<a href="http://www.etish.org/">http://www.etish.org/</a>	Solutions Member
45	ETRI (Electronics & Telecommunications Research Institute)	<a href="http://www.etri.re.kr/eng/">http://www.etri.re.kr/eng/</a>	Associate Member
46	Excelsior LLC	<a href="http://www.excelsior-usa.com/">http://www.excelsior-usa.com/</a>	Solutions Member
47	Exist Global (formerly DevZuz)	<a href="http://www.exist.com/">http://www.exist.com/</a>	Strategic Developer
48	Fraunhofer Institute for Open Communication Systems (FOKUS)	<a href="http://www.fokus.fraunhofer.de/">http://www.fokus.fraunhofer.de/</a>	Associate Member
49	Freescale Semiconductor	<a href="http://www.freescale.com/">http://www.freescale.com/</a>	Solutions Member
50	froglogic GmbH	<a href="http://www.froglogic.com/">http://www.froglogic.com/</a>	Solutions Member
51	Fujitsu Limited	<a href="http://www.fujitsu.com/">http://www.fujitsu.com/</a>	Solutions Member
52	Genuitec, LLC	<a href="http://www.genuitec.com/">http://www.genuitec.com/</a>	Strategic Developer
53	Gerhardt Informatics Kft.	<a href="http://www.gerhardtinformatics.com/">http://www.gerhardtinformatics.com/</a>	Solutions Member
54	Google Inc.	<a href="http://www.google.com/">http://www.google.com/</a>	Solutions Member
55	Hewlett-Packard Company (HP)	<a href="http://www.hp.com/">http://www.hp.com/</a>	Solutions Member
56	Hitachi, Ltd., Software Division	<a href="http://www.hitachi.com/">http://www.hitachi.com/</a>	Solutions Member
57	IBM	<a href="http://www.ibm.com/">http://www.ibm.com/</a>	Strategic Developer
58	IBS AB	<a href="http://www.ibs.net/">http://www.ibs.net/</a>	Solutions Member
59	IDG Japan	<a href="http://www.idg.co.jp/english/">http://www.idg.co.jp/english/</a>	Associate Member
60	IFS Institut fur Software	<a href="http://ifs.hsr.ch/">http://ifs.hsr.ch/</a>	Associate Member
61	ILOG	<a href="http://www.ilog.com/">http://www.ilog.com/</a>	Solutions Member
62	Ingres Corporation	<a href="http://www.ingres.com/">http://www.ingres.com/</a>	Solutions Member
63	Innoopract	<a href="http://www.innoopract.com/">http://www.innoopract.com/</a>	Strategic Developer
64	Innovations Softwaretechnologie GmbH	<a href="http://www.innovations-software.com/">http://www.innovations-software.com/</a>	Solutions Member
65	Instantiations, Inc.	<a href="http://www.instantiations.com/">http://www.instantiations.com/</a>	Solutions Member
66	instinctools GmbH	<a href="http://www.instinctools.com/">http://www.instinctools.com/</a>	Solutions Member
67	Intalio Inc.	<a href="http://www.intalio.com/">http://www.intalio.com/</a>	Solutions Member

68	Intel Corporation	<a href="http://www.intel.com/cd/software">http://www.intel.com/cd/software</a>	Strategic Consumer
69	Intervoice Inc.	<a href="http://www.intervoice.com/">http://www.intervoice.com/</a>	Solutions Member
70	IONA Technologies	<a href="http://www.iona.com/">http://www.iona.com/</a>	Strategic Developer
71	itemis AG	<a href="http://www.itemis.com/">http://www.itemis.com/</a>	Strategic Developer
72	iWay Software	<a href="http://www.iwaysoftware.com/">http://www.iwaysoftware.com/</a>	Solutions Member
73	Kestral Computing P/L (formerly Jiva Medical)	<a href="http://www.kestral.com.au/">http://www.kestral.com.au/</a>	Solutions Member
74	Klocwork	<a href="http://www.klocwork.com/">http://www.klocwork.com/</a>	Solutions Member
75	KPIT Cummins Infosystems Ltd.	<a href="http://www.kpitcummins.com/">http://www.kpitcummins.com/</a>	Solutions Member
76	Krugle, Inc.	<a href="http://www.krugle.com/">http://www.krugle.com/</a>	Solutions Member
77	Laszlo Systems	<a href="http://www.laszlosystems.com/">http://www.laszlosystems.com/</a>	Solutions Member
78	Lombardi Software	<a href="http://www.lombardisoftware.com/">http://www.lombardisoftware.com/</a>	Solutions Member
79	LSI Corporation	<a href="http://www.lsi.com/">http://www.lsi.com/</a>	Solutions Member
80	LynuxWorks Inc.	<a href="http://www.lynuxworks.com/">http://www.lynuxworks.com/</a>	Solutions Member
81	Macraigor Systems LLC	<a href="http://www.macraigor.com/">http://www.macraigor.com/</a>	Solutions Member
82	MDS Technology Co., Ltd.	<a href="http://www.mdstec.com/">http://www.mdstec.com/</a>	Solutions Member
83	Mentor Graphics	<a href="http://www.mentor.com/">http://www.mentor.com/</a>	Solutions Member
84	Metafinanz	<a href="http://www.metafinanz.de/">http://www.metafinanz.de/</a>	Solutions Member
85	Micro Focus	<a href="http://www.microfocus.com/">http://www.microfocus.com/</a>	Solutions Member
86	MicroDoc Computersystems GmbH	<a href="http://www.microdoc.com/">http://www.microdoc.com/</a>	Solutions Member
87	Mind8 GmbH	<a href="http://www.mind8.com/">http://www.mind8.com/</a>	Solutions Member
88	Misys International Banking Systems	<a href="http://www.misysbanking.com/">http://www.misysbanking.com/</a>	Solutions Member
89	MKS Inc.	<a href="http://www.mks.com/">http://www.mks.com/</a>	Solutions Member
90	MontaVista	<a href="http://www.mvista.com/">http://www.mvista.com/</a>	Solutions Member
91	Motorola	<a href="http://www.motorola.com/">http://www.motorola.com/</a>	Strategic Developer
92	MySQL	<a href="http://www.mysql.com/">http://www.mysql.com/</a>	Solutions Member
93	NEC	<a href="http://www.nec.com/">http://www.nec.com/</a>	Solutions Member
94	NetApp	<a href="http://www.netapp.com/">http://www.netapp.com/</a>	Solutions Member
95	Nexaweb Technologies	<a href="http://www.nexaweb.com/">http://www.nexaweb.com/</a>	Solutions Member
96	nexB Inc.	<a href="http://www.nexb.com/">http://www.nexb.com/</a>	Solutions Member
97	Nokia	<a href="http://www.nokia.com/">http://www.nokia.com/</a>	Strategic Developer
98	Novell	<a href="http://www.novell.com/">http://www.novell.com/</a>	Solutions Member
99	Obeo	<a href="http://www.obeo.fr/">http://www.obeo.fr/</a>	Strategic Developer
100	Object Management Group	<a href="http://www.omg.org/">http://www.omg.org/</a>	Associate Member
101	OC Systems	<a href="http://www.ocsystems.com/">http://www.ocsystems.com/</a>	Solutions Member
102	Open Source Lab	<a href="http://osuosl.org/">http://osuosl.org/</a>	Associate Member
103	Open Systems Publishing	<a href="http://www.opensystems-publishing.com/">http://www.opensystems-publishing.com/</a>	Associate Member
104	OpenMake Software Inc.	<a href="http://www.openmakesoftware.com/">http://www.openmakesoftware.com/</a>	Solutions Member
105	OpenMethods LLC	<a href="http://openmethods.com/">http://openmethods.com/</a>	Strategic Developer
106	Oracle	<a href="http://www.oracle.com/">http://www.oracle.com/</a>	Strategic Developer
107	OSGi Alliance, Inc.	<a href="http://www.osgi.org/">http://www.osgi.org/</a>	Associate Member

108	OW2	<a href="http://www.objectweb.org/">http://www.objectweb.org/</a>	Associate Member
109	Palamida, Inc	<a href="http://www.palamida.com/">http://www.palamida.com/</a>	Solutions Member
110	Parasoft	<a href="http://www.parasoft.com/">http://www.parasoft.com/</a>	Solutions Member
111	Perforce	<a href="http://www.perforce.com/">http://www.perforce.com/</a>	Solutions Member
112	Piterion	<a href="http://www.piterion.com/">http://www.piterion.com/</a>	Solutions Member
113	Polarion Software	<a href="http://www.polarion.com/">http://www.polarion.com/</a>	Solutions Member
114	Progress Software	<a href="http://www.progress.com/">http://www.progress.com/</a>	Solutions Member
115	Prosyst Software	<a href="http://www.prosyst.com/">http://www.prosyst.com/</a>	Solutions Member
116	Protecode Inc.	<a href="http://www.protecode.com/">http://www.protecode.com/</a>	Solutions Member
117	Purple Scout AB	<a href="http://www.purplescout.se/">http://www.purplescout.se/</a>	Solutions Member
118	Puzzle ITC GmbH	<a href="http://www.puzzle.ch/">http://www.puzzle.ch/</a>	Solutions Member
119	QNX Software Systems Co.	<a href="http://www.qnx.com/">http://www.qnx.com/</a>	Solutions Member
120	Quest Software, Inc.	<a href="http://www.quest.com/">http://www.quest.com/</a>	Solutions Member
121	Red Hat, Inc.	<a href="http://www.redhat.com/">http://www.redhat.com/</a>	Solutions Member
122	Redmond Media Group	<a href="http://redmondmediagroup.com/">http://redmondmediagroup.com/</a>	Associate Member
123	Remain BV	<a href="http://www.remain.nl/">http://www.remain.nl/</a>	Solutions Member
124	Replay Solutions	<a href="http://www.replaysolutions.com/">http://www.replaysolutions.com/</a>	Solutions Member
125	Research In Motion	<a href="http://www.rim.com/">http://www.rim.com/</a>	Solutions Member
126	RTC Group Inc.	<a href="http://www.rtcgroup.com/">http://www.rtcgroup.com/</a>	Associate Member
127	SalesForce.com	<a href="http://www.salesforce.com/">http://www.salesforce.com/</a>	Solutions Member
128	SAP AG	<a href="http://www.sap.com/">http://www.sap.com/</a>	Strategic Consumer
129	SAS	<a href="http://www.sas.com/">http://www.sas.com/</a>	Solutions Member
130	Serena Software, Inc.	<a href="http://www.serena.com/">http://www.serena.com/</a>	Solutions Member
131	Siemens	<a href="http://www.siemens.com/">http://www.siemens.com/</a>	Solutions Member
132	SIGS Datacom	<a href="http://www.sigs-datacom.de/">http://www.sigs-datacom.de/</a>	Associate Member
133	Skyway Software, Inc.	<a href="http://www.skywaysoftware.com/">http://www.skywaysoftware.com/</a>	Solutions Member
134	SlickEdit Inc.	<a href="http://www.slickedit.com/">http://www.slickedit.com/</a>	Solutions Member
135	Social Physics	<a href="http://www.socialphysics.org/">http://www.socialphysics.org/</a>	Associate Member
136	Software & Support Verlag	<a href="http://software-support.biz/">http://software-support.biz/</a>	Associate Member
137	Software AG	<a href="http://www.softwareag.com/">http://www.softwareag.com/</a>	Solutions Member
138	Sony Ericsson	<a href="http://www.sonyericsson.com/">http://www.sonyericsson.com/</a>	Solutions Member
139	Sopera GmbH	<a href="http://www.sopera.de/en/">http://www.sopera.de/en/</a>	Strategic Developer
140	Soyatec	<a href="http://www.soyatec.com/">http://www.soyatec.com/</a>	Solutions Member
141	SpikeSource Inc.	<a href="http://www.spikesource.com/">http://www.spikesource.com/</a>	Solutions Member
142	SpringSource, Inc	<a href="http://www.springsource.com/">http://www.springsource.com/</a>	Solutions Member
143	Standards for Technology in Automotive Retail	<a href="http://www.starstandard.org/">http://www.starstandard.org/</a>	Associate Member
144	STMicroelectronics	<a href="http://www.st.com/">http://www.st.com/</a>	Solutions Member
145	Sybase	<a href="http://www.sybase.com/">http://www.sybase.com/</a>	Strategic Developer
146	Symbian	<a href="http://www.symbian.com/">http://www.symbian.com/</a>	Solutions Member
147	Teamprise	<a href="http://www.teamprise.com/">http://www.teamprise.com/</a>	Solutions Member
148	Tensilica	<a href="http://www.tensilica.com/">http://www.tensilica.com/</a>	Solutions Member
149	Texas Instruments	<a href="http://www.ti.com/">http://www.ti.com/</a>	Solutions Member
150	Thales Group	<a href="http://www.thalesgroup.com/">http://www.thalesgroup.com/</a>	Solutions Member
151	The Medical Banking Project	<a href="http://www.mbproject.org/">http://www.mbproject.org/</a>	Associate Member

152	The RCP Company	<a href="http://rcp-company.com/">http://rcp-company.com/</a>	Solutions Member
153	TIBCO	<a href="http://www.tibco.com/">http://www.tibco.com/</a>	Solutions Member
154	TietoEnator Telecom R&D	<a href="http://www.tietoenator.com/">http://www.tietoenator.com/</a>	Solutions Member
155	TimeSys Corporation	<a href="http://www.timesys.com/">http://www.timesys.com/</a>	Solutions Member
156	Toshiba Corporation	<a href="http://www.toshiba.com/">http://www.toshiba.com/</a>	Solutions Member
157	VirtualLogix, Inc.	<a href="http://www.virtuallogix.com/">http://www.virtuallogix.com/</a>	Solutions Member
158	Virtutech, Inc.	<a href="http://www.virtutech.com/">http://www.virtutech.com/</a>	Solutions Member
159	Webtide	<a href="http://www.webtide.com/">http://www.webtide.com/</a>	Solutions Member
160	Weigle Wilczek GmbH	<a href="http://www.weiglewilczek.com/">http://www.weiglewilczek.com/</a>	Solutions Member
161	Wind River	<a href="http://www.windriver.com/">http://www.windriver.com/</a>	Strategic Developer
162	Zend Technologies	<a href="http://www.zend.com/en/">http://www.zend.com/en/</a>	Strategic Developer
163	Zensar Technologies, Ltd.	<a href="http://www.zensar.com/">http://www.zensar.com/</a>	Solutions Member

## 7.2. Appendix B: Descriptive statistics for the variables representing keywords

(sample size = 163)

	<b>Variable</b>	<b>Min</b>	<b>Max</b>	<b>Mean</b>	<b>Std. Deviation</b>	<b>Skewness</b>	<b>Kurtosis</b>
1	"associate member"	0	4.965	0.082	0.467	8.403	79.193
2	"based on" OR based-on	0	70.476	11.871	9.964	2.260	8.471
3	"community projects" OR "community project"	0	1.762	0.040	0.202	7.103	53.390
4	"cost reduction" OR "reduction of costs" OR "lower cost"	0	3.524	0.594	0.811	1.645	2.243
5	"Eclipse Foundation"	0	100.000	4.625	13.608	4.540	22.936
6	"Eclipse Public License" OR epl	0	28.235	0.520	2.668	8.550	80.795
7	"integration platform"	0	5.429	0.247	0.781	4.845	26.006
8	"Mike Milinkovich"	0	29.771	0.548	2.750	8.542	83.040
9	"open source"	0	95.541	12.436	19.730	2.234	4.793
10	"open source solutions"	0	39.886	0.563	3.336	10.489	121.234
11	"open technology"	0	13.333	0.124	1.056	12.256	153.810
12	"seamless integration"	0	6.015	0.368	0.691	4.505	29.359
13	"strategic member"	0	9.615	0.169	1.007	7.099	55.722
14	collaboration OR collaborate	0	49.220	4.770	6.929	3.247	13.934

15	commercially OR commercial	0	57.274	6.637	8.970	2.610	8.127
16	community	0	138.806	14.435	20.868	2.704	9.398
17	consistent OR consistently OR consistency	0	18.056	2.903	3.235	2.102	5.612
18	creating OR producing	0	52.381	7.495	8.046	2.455	8.273
19	customize OR customized	0	86.538	3.454	9.167	7.093	56.622
20	Eclipse	0	100.000	16.901	23.796	1.774	2.403
21	ecosystem OR ecosystems	0	51.806	1.852	4.985	6.990	63.516
22	embracing OR embrace OR embraces	0	8.105	0.564	1.069	3.858	19.373
23	equinox OR swordfish OR smila OR riena	0	28.235	0.843	3.152	5.990	42.650
24	extensibility OR extensible	0	37.460	1.806	4.351	6.167	44.492
25	flexibility OR flexible	0	33.755	6.200	5.955	1.724	4.009
26	framework OR frameworks	0	55.769	9.051	10.089	2.235	5.933
27	gef OR emf OR gmf OR emft OR gmt OR m2m	0	82.692	2.358	10.149	6.363	43.248
28	ide OR "development environment"	0	54.894	5.490	9.041	2.950	9.865
29	incorporating OR incorporate	0	9.037	1.436	1.738	1.941	4.098
30	invest OR investing OR investment	0	32.946	4.349	5.430	2.234	6.739

31	java OR ee OR j2ee OR j2se OR jdk OR jre OR jsf OR "java server faces"	0	1171.429	25.774	92.601	11.842	147.032
32	jdt OR cdt OR wtp OR dtp OR pde OR tptp	0	32.061	1.079	3.674	5.791	39.050
33	performance	0	65.753	15.660	14.401	1.146	0.906
34	perspective	0	27.778	2.382	3.800	4.075	20.772
35	platform	0	96.536	18.963	17.360	1.626	3.152
36	plugins OR plug-ins OR plugin or plug- in	0	86.538	3.410	8.740	6.224	51.894
37	quickly OR quick OR fast OR faster	0	78.846	10.906	10.634	2.386	10.117
38	rcp OR ercp OR babel OR ide	0	90.385	6.570	13.064	3.724	16.299
39	reliable OR reliability	0	37.248	4.856	5.984	2.437	7.703
40	scalable OR scalability	0	41.096	3.725	6.179	3.648	16.181
41	standardized OR standards- based OR "standards based" OR "open standards" OR standardization	0	14.237	2.080	2.754	2.049	4.480
42	strategic OR strategy	0	70.261	7.353	9.660	3.223	14.615
43	tool platform	0	17.647	1.927	2.855	3.019	10.771

### **7.3. Appendix C: Factor Analysis Techniques**

#### **7.3.1. Introduction to Factor Analysis**

Costello & Osborne (2005) define factor analysis as a complex procedure with few absolute guidelines and many different options. Many of these options have little or no explanation of their advantages and disadvantages, let alone when and why they should be used (Costello & Osborne, 2005; Darlington, 2007). This lack of documentation and strict guidelines has led to factor analysis being branded as unscientific, and has been marred with irreproducible results. In simple terms, factor analysis is used in order to discover patterns in the relationship among a number of variables. It specifically seeks to explain these relationships using a much smaller set of variables called factors.

“A successful factor analysis will provide answers to the following 4 questions:

1. How many different factors are needed to explain the pattern of relationships among these variables?
2. What is the nature of those factors?
3. How well do the hypothesized factors explain the observed data?
4. How much purely random or unique variance does each observed variable include?”(Darlington, 2007)

Factor analysis’ objectivity has come in to question due to the number of different results generated on the same data set by a number of different researchers. This does not mean that factor analysis is not objective however. Researchers will make trade-offs differently as they choose between simplicity (fewer factors) and accuracy (more factors). They may also differ in how they define their factors. Darlington (2007) points out that this will lead to multiple sets of correct solutions. This is very similar to algorithm development. When developing a particular algorithm, a software engineer must make trade-offs between

simplicity and performance. Although five different engineers will almost certainly develop five differing algorithms, it is also quite certain that all five will be correct.

#### **7.3.1.1. Factor Analysis' Notorious Past**

During the 1950's, factor analysis was positioned not as one of the data analysis techniques available, but as the only data analysis technique to use (Darlington, 2007). This overzealous promotion by a number of researchers led to the technique being over-used and misused. One particular incorrect use was trying to qualify very broad sets of data, pushing the technique further than it is able to return accurate results for. These factors, along with the complexity and lack of strict guidelines have led to factor analysis being mistrusted in some circumstances (Darlington, 2007). It is important to lay out a concise method indicating how factor analysis will be used in order to ensure correct results, and a defensible data set.

#### **7.3.2. Some Best Practices**

As mentioned before, factor analysis involves a multi-step method that allows the researcher to choose from a number of different techniques in each step. This section will outline some of the best practices that have been suggested by Costello & Osborne (2005) and Darlington (2007).

### **7.3.2.1. Principal Component Analysis vs. Factor Analysis**

Although it is the default extraction in most of the popular statistics packages, principal component analysis (PCA) is not often the best method to use (Costello & Osborne, 2005; Darlington, 2007). It is the most popular extraction technique to use however it is not a method of factor analysis using the classical definition. There is also a fair amount of disagreement about when PCA should and should not be used (Costello & Osborne, 2005). Because of this, Costello and Osborne recommend factor analysis over PCA. PCA is a data reduction method, made popular when computing power was expensive. In PCA there is no discrimination between unique and shared variance. Factor analysis on the other hand differentiates shared variance from unique variance and error variance in revealing the data sets underlying structure (Costello & Osborne, 2005). Factor analysis will result in the same solution as PCA, while at the same time avoiding the increase of variance by making use of a more complex algorithm which today's computers can handle with ease.

### **7.3.2.2. Factor Extraction Methods**

Costello & Osborne (2005) found that most statistics software packages have 6 or more factor extraction methods (PCA, unweighted least squares, generalized squares, maximum likelihood, principal axis factoring, alpha factoring and image factoring). This leads to confusion and usually to the use of PCA (which, as I have already mentioned is not usually the best choice). Costello and Osborne have outlined best practices for choosing an extraction method as well.

They recommend that if the data set is normal, maximum likelihood is the best choice. They refer to Fabrigar, Wegener, MacCallum & Strahan's (1999) reason for this choice because "[maximum likelihood] allows for the computation of a wide range of indexes of the goodness of fit of the model [and] permits statistical significance testing of factor loadings and correlations among factors and the computation of confidence intervals." When the data is not normal, Costello and Osborne recommend (once again referring to Fabrigar et al., 1999) the principal factor method (also referred to as principal axis factors).

#### **7.3.2.3. Deciding on the Number of Factors to Retain**

Choosing the number of factors is an important step in factor analysis. The higher the number of factors, the more complex the results will be to see, however, both Costello & Osborne (2005) and Darlington (2007) point out that it is sometimes necessary to use a larger set of factors in order to accurately reflect the relationships underlying the data set. It is important that the correct number of factors be retained for rotation since both over extraction and under extraction will lead to incorrect results. Most software packages use a terribly inaccurate method of selecting factors to retain: keeping all factors with eigenvalues greater than 1. Costello and Osborne state that this is inaccurate and will often lead to incorrect results. They recommend the scree test, Velicer's MAP criteria and parallel analysis. However, since the scree test tends to be the most widely available of the three it is their recommended choice. In the scree test, the graph of eigenvalues is examined and the number of data points above the "break" in the graph becomes the number of factors to retain. After rotation if the loading tables do not clearly show factors

then there may be a problem with either the number of factors or the data itself.

Removing the problematic items and re-running the analysis can help, but this can damage the data's integrity (Costello & Osborne, 2005). If, after multiple test runs the data does not take shape, the data may need to be purged and new data collected. In Darlington (2007) it is recommended that the scree test be used. As an alternate Darlington (2007) recommends completing a complete factor analysis with various numbers of factors and selecting the one that has the most well-defined structure.

Although more time and resource consuming, this will lead to the best results. From here on this will be referred to as the Darlington method for factor retention.

#### **7.3.2.4. Choosing Sample Size**

Although sample size tends to be a function of the strength of the data a good rule of thumb is that it is always better to have more data (Costello & Osborne, 2005; Darlington, 2007).