

**Examining governance of
open source software foundations**

by

Ludovico Pratico

A thesis submitted to the Faculty of Graduate and Postdoctoral Affairs
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

September 2012

© Copyright 2012 Ludovico Pratico



Library and Archives
Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

ISBN: 978-0-494-93512-5

Our file Notre référence

ISBN: 978-0-494-93512-5

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.

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

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

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

Canada

ABSTRACT

Artificial neural network analysis is used to analyse the content of the bylaws of six open source foundations for the purpose of identifying power structures. Members who contribute to open source foundations and use their products and services will find the results useful. Results of the research suggest that: i) the actions of an open source software foundation are centered around one of three groups: Members, Chairman/President/Executive Director, and Board of Directors; ii) only in one of the six foundations is the Board of Directors responsible for both the community and the product; and iii) artificial neural network analysis of the content of bylaws provides unbiased insights of the power structure of open source software foundations.

ACKNOWLEDGMENTS

Throughout this journey, there have been many individuals who have encouraged me along, who all merit an expression of gratitude.

First and foremost, I would like to express my sincerest heartfelt thanks and gratitude to, my advisor, Prof. Tony Bailetti. Without his relentlessly enthusiasm in providing feedback, encouragement and insight, along with him readily available to respond to questions, this work would have been a more difficult journey.

I am thankful to Prof. Steve Muegge, Prof. Michael Weiss and Prof. Mika Westerlund who provided me with valuable insight on where I should look to improve my research.

A big thank you goes out to my colleagues in the TIM program who cheered me on throughout the program.

Lastly, I would like to thank my family for their continued support and encouragement.

TABLE OF CONTENTS

| | |
|---|------|
| ABSTRACT..... | iii |
| TABLE OF CONTENTS..... | v |
| LIST OF TABLES..... | vii |
| LIST OF FIGURES | viii |
| LIST OF APPENDICES..... | ix |
| 1 INTRODUCTION..... | 10 |
| 1.1 Objective | 12 |
| 1.2 Deliverables..... | 12 |
| 1.3 Relevance | 13 |
| 1.4 Contributions..... | 14 |
| 1.5 Organization | 15 |
| 2 LITERATURE REVIEW | 16 |
| 2.1 Open Source Software Foundations..... | 16 |
| 2.2 OSSF Governance | 18 |
| 2.3 Stakeholder Influence..... | 20 |
| 2.4 Artificial Neural Network Analysis | 22 |
| 2.5 Lessons learnt from the literature..... | 25 |
| 3 RESEARCH METHOD | 26 |
| 3.1 Unit of analysis..... | 26 |
| 3.2 Study period | 26 |
| 3.3 Research method | 26 |
| 3.4 Research Steps..... | 28 |
| 3.4.1 Identify foundations..... | 28 |
| 3.4.2 Acquire data..... | 29 |
| 3.4.3 Analyse bylaws | 29 |
| 3.4.4 Identify concepts..... | 30 |
| 3.4.5 Categorise bylaws | 30 |
| 3.4.6 Categorise OSSF Mission Statements | 30 |
| 3.4.7 Generate insights..... | 30 |
| 3.5 Open Source Foundations Studied | 31 |
| 3.5.1 The Apache Software Foundation | 31 |
| 3.5.2 The Eclipse Foundation | 32 |
| 3.5.3 The GNOME Foundation | 32 |
| 3.5.4 The Plone Foundation..... | 33 |
| 3.5.5 The Python Software Foundation | 33 |
| 3.5.6 The Software in the Public Interest (SPI) Foundation..... | 34 |
| 3.5.7 Open source software foundations mission statements | 34 |

| | | |
|-------|---|----|
| 3.6 | Data Collected | 36 |
| 4 | RESULTS | 37 |
| 4.1 | Data Analysis | 37 |
| 4.1.1 | Apache Software Foundation..... | 37 |
| 4.1.2 | The Eclipse Foundation | 41 |
| 4.1.3 | The GNOME Foundation | 45 |
| 4.1.4 | Plone Foundation | 49 |
| 4.1.5 | Python Foundation..... | 53 |
| 4.1.6 | Software in the Public Interest Foundation..... | 57 |
| 4.2 | Summary of results..... | 62 |
| 5 | DISCUSSION OF RESULTS | 65 |
| 5.1 | Categorisation of foundations using bylaws | 65 |
| 5.2 | Answering the research question..... | 66 |
| 5.3 | Categorisation of OSSF..... | 67 |
| 5.4 | Links to the literature | 67 |
| 5.5 | Anomalies in the research | 68 |
| 5.6 | Challenges and insights of using computer aided text analysis | 69 |
| 5.7 | Insights for stakeholders..... | 71 |
| 6 | CONCLUSIONS, LIMITATIONS AND FUTURE RESEARCH..... | 73 |
| 6.1 | Conclusions | 73 |
| 6.2 | Research Limitations..... | 74 |
| 6.3 | Suggestions for Future Research..... | 75 |

LIST OF TABLES

| | |
|--|-----------|
| Table 1. Definitions of stakeholder attributes..... | 21 |
| Table 2. Influencing Strategies | 22 |
| Table 3. Research Steps. | 28 |
| Table 4. Open source foundations studied..... | 31 |
| Table 5. Mission statements characterisation | 35 |
| Table 6. Documents collected..... | 36 |
| Table 7. Mission statements characterisation based power centre | 67 |

LIST OF FIGURES

| | |
|--|----|
| Figure 1. ASF dendrogram analysis..... | 38 |
| Figure 2. ASF 2D Conceptual Map | 39 |
| Figure 3. ASF 3D Conceptual Map | 40 |
| Figure 4. Eclipse Foundation Dendrogram..... | 42 |
| Figure 5. Eclipse Foundation 2D Conceptual Map..... | 43 |
| Figure 6. Eclipse Foundation 3D Conceptual Map..... | 44 |
| Figure 7. GNOME Foundation Dendrogram | 46 |
| Figure 8. GNOME Foundation 2D Conceptual Map..... | 47 |
| Figure 9. GNOME Foundation 3D Conceptual Map..... | 48 |
| Figure 10. Plone Foundation Dendrogram..... | 50 |
| Figure 11 Plone Foundation 2D Conceptual Map | 51 |
| Figure 12. Plone Foundation 3D Conceptual Map | 52 |
| Figure 13. Python Foundation Dendrogram | 54 |
| Figure 14. Python Foundation 2D Conceptual Map | 55 |
| Figure 15. Python Foundation 3D Conceptual Map | 56 |
| Figure 16. SPI Foundation Dendrogram..... | 59 |
| Figure 17. SPI Foundation 2D Conceptual Map..... | 60 |
| Figure 18. SPI Foundation 3D Conceptual Map..... | 62 |
| Figure 19. Categorisation of foundations..... | 66 |

LIST OF APPENDICES

Appendix A. CATPAC II and ThoughtView Software 81

Appendix B. OSSF Mission Statements Provenance 84

1 INTRODUCTION

The number of open source software foundations (OSSFs) is steadily increasing. Riehle (2010) argues that there are good reasons for this. The motivations are economic in that it saves in development costs by spreading the cost across participating members, increases revenue generation through increased sales of complementary products, increases the addressable market by competing more effectively across “technology stacks” and as a common good, the firms involved increase their goodwill and general welfare. Typically, OSSFs create, enhance and support open source tools, frameworks and software systems that include operating systems, productivity software and content management systems. In addition, the foundations are keystones anchoring ecosystems that include companies that generate revenue by developing and commercialising products based on the outputs produced by the OSSFs.

In recent blog posts and related magazine articles, the question as to whether open source software foundations are trade associations or open-source communities, has been asked (Taft, 2009). One conclusion is that they can be both (Milinkovich, 2010). This leads to questions as to where the power lies in the governance of open source software foundations. Who is responsible for what?

In order to guide their operations and achieve their objectives, OSSFs develop governance policies in at least four areas:

1. Board of Directors composition

2. Foundation bylaws
3. Membership agreement
4. Intellectual property rights

Often, the policy areas are combined into one document referred to as bylaws.

This research attempts to discover whom of the Chairman/President/Executive Director, Board of Directors or Members has the power in open source software foundations (OSSFs). For the purposes of this research, power is defined as the capability of one social actor to overcome a resistance in achieving a desired objective (Pfeffer, 1981).

Power has many sources that include delegated authority, social class, material resource, charisma, knowledge, expertise and so on (Pfeffer, 1981). In addition, French and Raven (1959) define five bases of power: reward, coercive, legitimate, referent and expert. This research looks at legitimate power stemming from internalised values in one social actor, A, that another social actor, B, has a legitimate right to influence A, who is obliged to accept the influence of B (French & Raven, 1959).

Through the analysis of company involvement, governance and OSSF effectiveness, Xie (2008) studied their effect on OSSF effectiveness and profitability. Based on researcher knowledge and experience, Xie concluded there are three different power structures in

OSSFs, namely Merit, Merit Dominated and Sponsor Dominated. This research builds on the work carried out by Xie (2008), studying the power structures within OSSFs by analysing the governance documents. Discourse analysis (George & Bock, 2010) also known as content analysis (Neuendorf, 2002) is used to analyse the documents. By using computer aided text analysis (Neuendorf, 2002), the analysis is performed without researcher bias arising from personal experience and knowledge.

1.1 Objective

The objective of this research is to answer the question, “Where does the power lie in the governance of open source software foundations?”

The question is answered by examining the bylaws of not-for-profit, member-supported, open source software foundations that are the keystone organisations upon which open source software products are anchored on.

1.2 Deliverables

The research deliverables are:

- A categorisation of bylaws of open source software foundations
- Description of the links between the categories and the mission statements of open source foundations
- A procedure to generate perceptual maps of OSSF bylaws

1.3 Relevance

This research is relevant to top management teams of small and large companies, investors, consumers, government regulators, and personnel who work for economic development organisations.

The research is relevant to top management teams of small and large companies, and investors because it may help them better understand how OSSFs operate to achieve their objectives.

Government regulators interested in ensuring there is adequate competition in the marketplace will find this research relevant because it highlights who the foundations serve best, whether it is the communities they purport to serve or the entities that establish, operate and fund the OSSFs or the people who work for the OSSFs. In addition, governments interested in reducing their costs by using open source solutions and in encouraging development of new products based on open source solutions will find it relevant in making purchase decisions.

Economic development organisations, often faced on decisions as to where it is best to invest their limited resources, benefit from this research because they will get a better understanding of the impact the foundations have on the ecosystems they have spawned.

Academics will be interested in this research as it presents a method for doing discourse analysis (content analysis) of organisations' bylaws that reduces researchers' biases.

This research is relevant to consumers interested in adopting open source, standards based tools, which will be supported and maintained over the long term.

1.4 Contributions

This thesis makes at least three contributions. First, by making visible the governance structures that exists today, this research facilitates the design and implementation of governance mechanisms for individuals who wish to launch new open source foundations.

The second contribution this thesis makes is that it shows how perceptual maps can be applied to make sense of complex written documents reducing the biases that may be introduced by the experience and knowledge of researchers.

The third contribution of the thesis is that it identifies the differences in the governance structures that exist today in open source environments. This can assist stakeholders articulate their views and system of beliefs clarifying issues and spawning innovative ideas for future governance development and research.

1.5 Organization

The thesis is organized into six chapters. Chapter 1 is the introduction. Chapter 2 reviews the stakeholder theory and profiting from innovation literature streams, and identifies the lessons learned from reviewing them. Chapter 3 describes the research method. The Chapter 4 presents the results. The results of the research are discussed in Chapter 5. Chapter 6 provides conclusions, describes the research limitations, and identifies opportunities for future research.

2 LITERATURE REVIEW

This chapter is organised into five sections. Section 2.1 is a review of open source software foundation models and their effectiveness. Section 2.2 reviews the literature in open source software foundations governance. Section 2.3 reviews the literature stream relevant to stakeholder influence. In Section 2.4, the application of neural network analysis to solving management problems is reviewed. Section 2.5 provides the lessons learnt from the literature review.

2.1 Open Source Software Foundations

Fitzgerald (2006) argues that open source software (OSS) has been transformed to OSS 2.0, to a more commercial business, gradually putting an end to the proprietary model of software development. In addition, he notes that proprietary firms' are increasing their involvement in the open source and free software industry as collaborative participants rather than competing rivals.

A framework to characterise the OSS based on the development lifecycle, product domain, primary business strategy, product support and licensing is proposed by Fitzgerald (2006). He argues that OSS 2.0 has moved from infrastructure to more visible information systems applications. These applications have a development lifecycle that is

more structured. It uses rigorous project management approaches and developers are paid. They extend the value-added service-enabling and loss-leader business strategies of FLOSS to include cost reduction, new feature introduction, multiple licences and leveraging the open source brand. Support has moved from email support groups to purchased professional support.

Jin, Robey and Boudreau (2007) see user support as a significant distinction between OSS and proprietary software in that in OSS, the developer is no longer liable for after-sales support. This results in a shift in the software technical support from being an additional cost to the developer company to being a separate consultation and training industry.

Dahlander and Wallin (2006) analyze firms' attempts to access complementary innovation assets of OSSFs through sponsored individuals within OSSFs. The authors' starting point is that the use of firm sponsorship is as frequent as it is the primary means to reach complementary assets of an open source community, otherwise beyond reach, and is one of the key determinants of a firm's power to appropriate innovations (Teece, 1986, 2006). Dahlander and Wallin's results indicate that despite the more intense activities and resource advantages of the sponsored individuals compared to the volunteering hobbyists in the community, these individuals have no significant effect in the community dynamics and interaction in the form of stimulating and/or leading

debates, or through gaining an important role in the community. On the other hand, firms, to some extent, can influence the direction of the development process of OSSFs significantly through participation and putting to use the resources into the process. The study also contributes to the evolution of innovation environments, the key drivers of open innovation and their affects on software development industry, and the firms' incentives and means for improving their innovative power.

2.2 OSSF Governance

An overview of governance mechanisms is provided by de Laat (2007). He discusses three types of governance that have been uncovered in the studies of open source software development. They are spontaneous governance, internal governance and “governance towards outside parties”. Spontaneous governance is characterised by communities of volunteers who enjoy the intellectual stimulus, have a desire to learn and improve their skillset, need the code created for their current professional employment or personal use. These communities are cross-institutional boundaries are self-directing with no formal control. Typically the de facto leaders are the 20% that produce 80% of the code.

The second governance method is related to projects that use explicit and formal tools to co-ordinate and control OSS projects. These projects are deemed to use internal

governance. De Laat (2007) argues that internal governance is characterised by six groups of tools: modularization, division of roles, delegation of decision-making, training and indoctrination, formalization and autocracy/democracy.

The third governance method is “governance towards outside parties”. This form of external governance is a result of outside parties, such as firms, governments and non-governmental organisations taking an interest in the benefits of open source software. In order to deal with the challenges associated with creating software in the commons and the threats from patent infringement, a “legal shell” (deLaat, 2007), was created around the OSS development.

Xie (2008) uses the term governance structures to refer to “who participates in the decision making”. Xie concludes that there are three types of governance structures, namely (i) Merit, (ii) Merit Dominated and (iii) Sponsor Dominated. In foundations with Merit governance structures, all members are merit members with full voting rights. In foundations with Merit Dominated governance structures, merit members are the majority making it difficult for sponsor members to affect the outcomes. In foundations with Sponsor Dominated governance structures sponsored members are typically company employees and would have a greater say in decisions.

O’Mahoney (2007) discusses what it means to be community managed. From the research on 4 large and mature OSS communities, he identified 5 principles for the

community managed governance model. These are: (i) Independence of any one sponsor; (ii) Pluralism in diversity of contributors, management of conflict, determination of leadership; (iii) Representation where contributing members can be represented in all community decisions; (iv) Decentralised decision making such as how contributors gain access to decision-making structures; and (v) Autonomous participation in that all contributors are welcomed and members contribute on their own terms.

2.3 Stakeholder Influence

As reported in Mitchell, Agle and Wood (1997), Freeman (1984) defined stakeholders as: “A stakeholder in an organization is (by definition) any group or individual who can affect or is affected by the achievement of the organization’s objectives.” According to Mitchell et al.(1997), this definition is ambiguous in that it encompasses just about anyone. To reliably identify stakeholders, they developed the concept of stakeholder salience, defined as “the degree to which managers give priority to competing stakeholder claims” and “to whom and to what managers pay attention.” Then to describe saliency, they propose three stakeholder attributes: Power, Legitimacy, and Urgency. Table 1 provides definitions for these three stakeholder attributes (Mitchell et al., 1997).

| Attribute | Definition |
|------------------|--|
| Power | The ability of one actor to persuade another actor to willingly or unwillingly take an action to the benefit of the first actor |
| Legitimacy | A generalised perception or assumption that the actions of an entity are desirable, proper or appropriate within some socially constructed system of norms, values and beliefs |
| Urgency | The degree to which stakeholder claims call for immediate actions |

Table 1. Definitions of stakeholder attributes

Using the theory of resource dependence as a framework, Frooman (1999) addresses the question, “How can external entities influence an organization’s behaviour?” To answer the question, Frooman examined the relationship between the firm and the stakeholder by looking at the dependence of the firm on the stakeholder’s resource and by studying the strategies the stakeholder had to influence the firm based on the relationship. A firm can be either dependent or independent of a stakeholder’s resource. If the firm has no dependence on the stakeholder’s resource and the stakeholder has no dependence on the firm, then there is a low interdependence between the firm and the stakeholder. If the firm has high dependence on the stakeholder’s resource and the stakeholder has no dependence on the firm, then the stakeholder has the power. If the firm has no dependence on the stakeholder’s resource and the stakeholder has dependence on the firm, then the firm has the power. Finally, if both the firm and the stakeholder have a dependence on each other, then it is a high interdependence relationship.

To influence the firm, a stakeholder has two options: (i) resource control strategies such as withholding the resource, and (ii) pathways of influence such as having direct influence on the firm or indirect influence through a third party.

Frooman (1999) concluded that stakeholders have four types of influencing strategies based on directly and indirectly withholding the resource and direct and indirect usage of the resource. Table 2 identifies the four types of influencing strategies.

| Resource relationship (Between the firm and stakeholder) | Stakeholder influencing strategy |
|---|---|
| Low interdependency | Indirect withholding of resource |
| High interdependency | Direct usage |
| Firm power | Indirect usage |
| Stakeholder power | Direct withholding |

Table 2. Influencing Strategies

2.4 Artificial Neural Network Analysis

An artificial neural network (ANN) is capable of discovering patterns in data by identifying frequencies and clustering of recurring concepts. It has been used for qualitative research in various disciplines including business and sociology.

Allen (2005) used ANN analysis to study sustainable forest management in British Columbia, Canada, from the perspective of First Nations peoples. The study used data acquired through personal interviews and oral history from forest-dependent First Nations communities. The analysis indicated that tribal elders were concerned with forest harvesting and loss of territory, forest health in terms of invasive species devastating the forest, access to wild life food sources and protection of old growth territories as habitats for fisheries and wildlife. The study also concluded the use of ANN analysis and concept mapping to text-based data can assist stakeholders articulate their views and system of beliefs.

Ryan (1998) used ANN analysis to study saltwater crocodiles as tourist attractions in the Northern Territory of Australia. A model was proposed to determine what attributes of a crocodile attracted tourists. For the study, 50 respondents were interviewed. From the ANN analysis, it was discovered that the model was incomplete as it failed to capture the ambivalence tourists felt about crocodiles. The research also revealed that a smaller number of respondents felt that although crocodile based attractions are interesting, they had too much of a circus component to them. This aspect of ANN analysis is useful in uncovering concepts that are not immediately obvious.

Evans et al. (2010) used ANN analysis to try uncovering patterns in multilingual texts. Chinese, English, Farsi, Hindi and Korean translations of the United Nations' Universal Declaration of Human Rights (UNUDHR) were analysed. It was surmised that since

ANN analysis is independent of grammar and the actual meaning of words, but dependent on their coded representation (UNICODE in this case), the concepts in the document should cluster the same way for all five translations. However, there was no clear finding from the analysis. They discovered that although the number of concept clusters was similar for all languages, the languages clustered differently. In the end, the authors concluded that the UNUDHR was not an ideal document for the analysis as it is short and written in formal language rather than narrative language.

Samkin and Schneider (2008) used ANN analysis to investigate whether or not there were notable differences between the concepts of control in two accounting standards, FRS (2001) and NZ IAS 27 (2004), used in New Zealand. In this investigation, the researchers extracted the relevant section from the accounting handbooks. The text was pre-processed manually to ensure uniform spelling, consistent treatment of synonyms, homonyms and words with negative usage not key to the text. From the analysis of the subset of the accounting handbooks, the researchers discovered that the concept of control in the older accounting practice was more detailed than the newer one. Furthermore, the researchers concluded that ANN analysis is suitable to analyse management discussion letters in annual reports, compare management letters over the years to identify recurring themes, among other uses related to management practices.

2.5 Lessons learnt from the literature

At least four lessons were learned from the literature.

1. Independence, pluralism, representation, decentralised decision making and autonomous participation are the five principles that characterise member-based organisations.
2. Due to the rapid growth of open source software foundations and their emerging governance structures, along with the diversity of stakeholders, there is a need for both qualitative and quantitative investigation that enhance governance of OSSFs.
3. OSSFs can benefit from adopting a broad view of the theory of stakeholder identification and salience that concentrates on maximising all stakeholders' interests.
4. Artificial neural network analysis is an acceptable method of performing content analysis of text-based documents, to generate new ideas from the text with minimal researcher bias.

3 RESEARCH METHOD

This chapter describes the research method and data collected. It is organized into four sections. Section 3.1 defines the unit of analysis. Section 3.2 defines the study period and data sample. Section 3.3 describes the research method. Section 3.4 describes the research steps. Section 3.5 provides brief descriptions of the foundations studied. Section 3.6 lists the provenance of the data collected.

3.1 Unit of analysis

The unit of analysis is the most recent bylaws of open source software foundations.

3.2 Study period

The period studied is two years, starting January 2009, ending December 2010. The bylaws were reviewed again in June 2012 to ensure there had been no changes since the studied period.

3.3 Research method

This research uses content analysis (Neuendorf, 2002), also called discourse analysis (George & Bock, 2010), to extract messages and ideas from text. Content analysis tools have been developed primarily for research in social sciences (George & Bock, 2010), but have also been used in various research applications that include organizational change and strategic management (George & Beck, 2010), and coding of student drawings and observations in videotaped studies (Stemler, 2001). Neuendorf (2002)

writes that one content analysis method suitable for text analysis is semantic network analysis. This method involves representing the text to be analysed as a network of objects by connecting frequently occurring words based on their co-occurrence. From the co-occurrence data, cluster analysis can be performed and by using multidimensional scaling analysis (MDS), visualisations of the co-occurrences can be created. The techniques lend themselves very well to computer aided text analysis. One such software application is CATPAC II TM (Wölfel, 1998).

CATPAC is a software implementation of a self-organising artificial neural network, optimised for analysing text (Wölfel, 1998). The software is used to establish patterns within text. It identifies the most important words and determines patterns of similarity based on how the words are used. It achieves this by running a scanning window through the text to create a multidimensional matrix where each words is assigned a weighting based on the distance from other words. From this matrix, ideas can be identified using conceptual maps and cluster diagrams. A more detailed discussion of CATPAC can be found in Appendix A.

3.4 Research Steps

Table 3 provides the steps required to complete this research.

| Step | Dominant activity | Outcome from activity |
|-------------------------|---|--|
| 1. Identify foundations | Identify foundations to examine | Foundations identified |
| 2. Acquire data | Gather the bylaws and convert to plain text files | Text version of bylaws document |
| 3. Analyse bylaws | Generate the visualisations of the bylaws using multidimensional scaling analysis | Visualisations of key players relative to the actions they guide |
| 4. Identify concepts | Identify concepts that can be used to categorise the governance structures | Concepts based on actions of key players |
| 5. Categorise bylaws | Categorise the bylaws based on concepts identified | A categorisation of the OSSFs based on bylaws |
| 6. Categorise OSSFs | Categorise OSSFs' mission statements based on the power centre | A categorisation of OSSFs based on mission statements |
| 7. Generate insights | Generate insights based on the analysis | Insights |

Table 3. Research Steps.

3.4.1 Identify foundations

This research uses the same sample used by Xie (2008). The criteria used by Xie was that the foundation in the sample must:

1. Have at least one active open source software project
2. Be incorporated in the United States
3. Have 501(C) tax exempt status in 2009 and 2010

This ensures that the foundations studied are active, not-for-profit in a common jurisdiction, operating under the same laws.

3.4.2 Acquire data

The data were the bylaws. The bylaw documents were downloaded from the websites of the foundations in the sample.

Inspection of the documents downloaded showed that, in general, the bylaws of the foundations incorporated the rules on the foundation's management structure, the bylaws and the details of the membership agreement.

The bylaws documents were in one of three formats: DOC, HTML or PDF. For analysis, the bylaw documents had to be converted to text (TXT). This conversion was achieved through the use of tools such as Microsoft Word and Adobe PDF Creator, to strip out formatting information.

3.4.3 Analyse bylaws

The following steps are undertaken to analyse the bylaw text:

1. Perform a word-count analysis to identify high-count words.
2. From this a list of words such as prepositions, conjunctions, and articles, create an exclusion list of high-count words that do not contribute to the analysis.
3. Repeat steps 2 & 3 if more non-contributing high-count words are identified.

4. Once all non-contributing words have been eliminated, create the dendrogram¹ (hierarchical cluster diagrams) and conceptual (MDS) maps using the CATPAC tool.
5. Identify the terms of interest

3.4.4 Identify concepts

Based on the terms of interest, develop a categorisation scheme for the foundation and the document analysed.

3.4.5 Categorise bylaws

Based on the analysis of the bylaws and where their power is centred, generate a categorisation of bylaws.

3.4.6 Categorise OSSF Mission Statements

The mission statements of the six foundations were organized based on their orientation for (i) community, (ii) product, or (iii) community and product.

3.4.7 Generate insights

Insights that were deemed to be of interest to the groups this research is relevant to were generated.

¹ Some authors have used the spelling *dendogram* for the hierarchical cluster diagram. The proper term is dendrogram from the from Greek *dendron* meaning "tree", *-gramma* meaning "drawing". (OED, 2012)

3.5 Open Source Foundations Studied

The bylaws of six open source software foundations were used in this study. Table 4 lists the foundations whose bylaws were examined and provides their Uniform Resource Locators (URLs).

| OSF | URL |
|-------------------|--|
| Apache Foundation | http://apache.org/ http://apache.org/foundation/ |
| Eclipse | http://www.eclipse.org/ |
| GNOME | http://www.gnome.org http://www.gnome.org/foundation/ |
| Plone | http://plone.org/ http://plone.org/foundation |
| Python | http://www.python.org/ http://www.python.org/psf/ |
| SPI | http://www.spi-inc.org/ http://www.spi-inc.org/corporate/ |

Table 4. Open source foundations studied

3.5.1 The Apache Software Foundation

The Apache Software Foundation (ASF) is a non-profit corporation established under the laws of the United States. In June 1999, the ASF was created from the Apache Group and incorporated in the US State of Delaware. The corporation was set up to support Apache software projects, responsible for over 60 projects, the most notable being the Apache HTTP Server and Open Office. The ASF is a decentralized community of

developers and distributes the software products it produces under the Apache License. Anyone can become a member, but one has to be nominated by another member. The Board of Directors is elected and any member can be elected to the Board.

The ASF is a US 501(c)(3) non-profit organization.

3.5.2 The Eclipse Foundation

The Eclipse Consortium was established in 2001 following IBM releasing the Eclipse project as open source software. The Consortium was initially formed with about 50 members. In May 2004, the Consortium formed the Eclipse Foundation. The Foundation has several levels of membership. The membership levels include Strategic, Enterprise, Solutions and Associate Members. The Strategic members are on the Board of Directors and have direct influence on the strategic direction.

The Eclipse Foundation is a US 501(c)(6) non-profit corporation.

3.5.3 The GNOME Foundation

The GNOME Foundation was established to further the goals of the GNOME project, a community of software developers, artists and active users of GNOME, a desktop environment and development platform. It also host various other projects related to GNOME. It is a non-profit organization based in Cambridge, Massachusetts, United States. All GNOME contributors can apply for membership and every member of the

Board of Directors is a contributing member of the community. The Board of Directors is elected every year from the membership.

The GNOME Foundation is a US 501(c)(3) non-profit organization.

3.5.4 The Plone Foundation

Alexander Limi, Alan Runyan, and Vidar Andersen started the Plone project in 1999 to support a free and open source content management system built on top of the Zope application server. The project quickly grew into a community requesting enhancements and additions. In response, in May 2004, the Plone Foundation was founded for the development, marketing, and protection of Plone. Anyone who has contributed significantly to the PLONE project can apply for membership. PLONE has an elected Board of Directors.

The Plone Foundation is a US 501(c)(3) non-profit organization.

3.5.5 The Python Software Foundation

The Python Software Foundation (PSF), formed on March 6, 2001 is a non-profit organization devoted to promote, protect, and advance the Python programming language, and “to support and facilitate the growth of a diverse and international community of Python programmers.” (PSF Mission Statement, 2012)

There are three classes of members in the PSF, namely nominated members, sponsor members and emeritus members. The Board of Directors is elected from the members.

The Python Software Foundation (PSF) is a US 501(c)(3) non-profit corporation.

3.5.6 The Software in the Public Interest (SPI) Foundation

The SPI Foundation, a non-profit organisation incorporated in June 1997, was founded to help organisations develop and distribute open hardware and software, acting as a fiscal sponsor to many free and open source projects. SPI has two types of memberships, contributing and non-contributing. Anyone can apply for membership, however to apply for contributing membership, one must sign-up for non-contributing status. Contributing memberships are open to all those who participate in the community and have voting rights. The Board of Directors is elected from contributing members.

The SPI Foundation is a US 501(c)(3) non-profit corporation.

3.5.7 Open source software foundations mission statements

Table 5 lists the mission statements of the open source software foundations studied. For clarity, the provenance of the mission statement is detailed in Appendix B.

In Table 5, the mission statements are categorised according to whether they are oriented to the community of users and developers or the product the OSSF sponsors.

| OSSF | Mission Statement | Orientation |
|-------------|--|-----------------------|
| Apache | The Apache Software foundation provides support for the Apache community of open-source projects. The Apache projects are characterised as collaborative, consensus based development process, an open and pragmatic license and a desire to create high quality software that leads the way in its field. | Community |
| Eclipse | The Eclipse Foundation was created in January 2004 as an independent not-for-profit corporation to act as the steward of the Eclipse community, to allow a vendor neutral and open, transparent community to be established around Eclipse. | Community |
| GNOME | The GNOME Foundation will work to further the goal of the GNOME project: to create a computing platform for use by the general public that is completely free software. | Product |
| Plone | The mission of the foundation is to protect and promote Plone. Our goals include: <ul style="list-style-type: none"> • Provide clear, neutral, and sustainable ownership of code, trademarks, and domains. • Provide a decision-making structure for essential community activities. • Ensure that, as Plone grows, it remains a level playing field. • Act as the voice of Plone for official announcements, press releases, and other communications. • Help create promotional material, interviews, speeches, and other activities to market Plone. | Product |
| Python | The mission of the Python Software Foundation is to promote, protect, and advance the Python programming language, and to support and facilitate the growth of a diverse and international community of Python programmers. | Community and Product |
| SPI | To create, form and establish an organization to formulate and provide software systems for use by the general public without charge and acts as a fiscal sponsor to many free and open source projects according to their website. | Product |

Table 5. Mission statements characterisation

3.6 Data Collected

Table 6 list the documents collected. The data was downloaded from its respective website during the month of December 2010.

| OSF | Document | URL |
|---------|----------|--|
| Apache | Bylaws | http://apache.org/foundation/bylaws.html |
| Eclipse | Bylaws | http://www.eclipse.org/org/documents/ |
| GNOME | Bylaws | http://www.gnome.org/foundation/governance/ http://www.gnome.org/wp-content/uploads/2011/11/bylaws.pdf |
| Plone | Bylaws | http://plone.org/foundation/materials/corporate http://plone.org/foundation/materials/corporate/Plone_Bylaws200809.doc/view |
| Python | Bylaws | http://www.python.org/psf/bylaws/ |
| SPI | Bylaws | http://www.spi-inc.org/corporate/ http://www.spi-inc.org/corporate/by-laws/ |

Table 6. Documents collected

The bylaws were downloaded from their respective websites and converted to text format. They were used as the input data to CATPAC, the artificial neural network software used to detect patterns in written text.

4 RESULTS

This chapter is organized into two sections. Section 4.1 contains the results of the artificial neural network analysis illustrated as hierarchical cluster analysis diagrams (dendrograms), two-dimensional and three-dimensional perceptual maps, also known as conceptual maps. Section 4.2 provides an overview of the results.

4.1 Data Analysis

The following subsections describe the output of the software for each of the foundations. Three figures are presented, the hierarchical cluster diagram or dendrogram, a two-dimensional conceptual map and a three-dimensional conceptual map.

4.1.1 Apache Software Foundation

This section describes the results of the analysis of the data for the Apache Software Foundation (ASF).

ASF Hierarchical Cluster Analysis (Dendrogram)

Figure 1 illustrates the dendrogram for the Apache Software Foundation. It illustrates a number of loosely tied together concepts. The only significant cluster is the first one that ties together the ACTIONS of the foundation to COMMITTEES involving a CHAIRMAN and MEMBERSHIP. No clear message can be extracted from the other clusters.

CENTROID METHOD

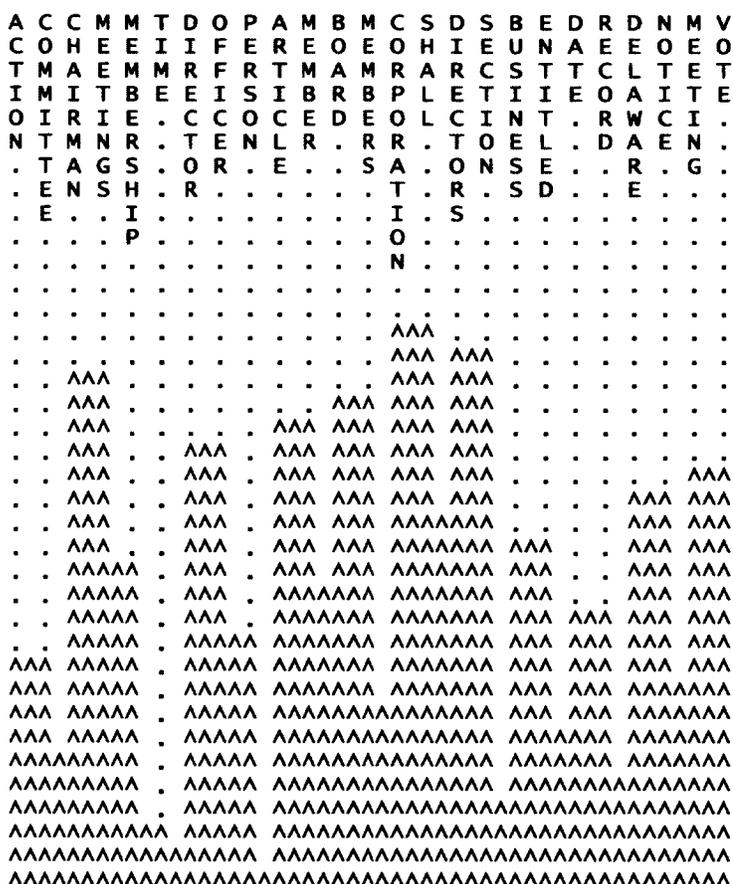


Figure 1. ASF dendrogram analysis

ASF Two Dimensional Conceptual Map

Figure 2, the 2-D conceptual map, shows the clustered data represented over 4 quadrants. This view leads to the same conclusions as the dendrograms. In Quadrant 1, we have the concepts of ACTION, CHAIRMAN, COMMITTEE and MEMBERSHIP all clustered together. Quadrant II indicates a close relationship between the MEMBERS and the BOARD. Quadrant III clusters the concepts of DIRECTORS, MEMBERS and the CORPORATION (Foundation). Lastly, Quadrant 4 indicates a close relationship between OFFICER and DIRECTOR.

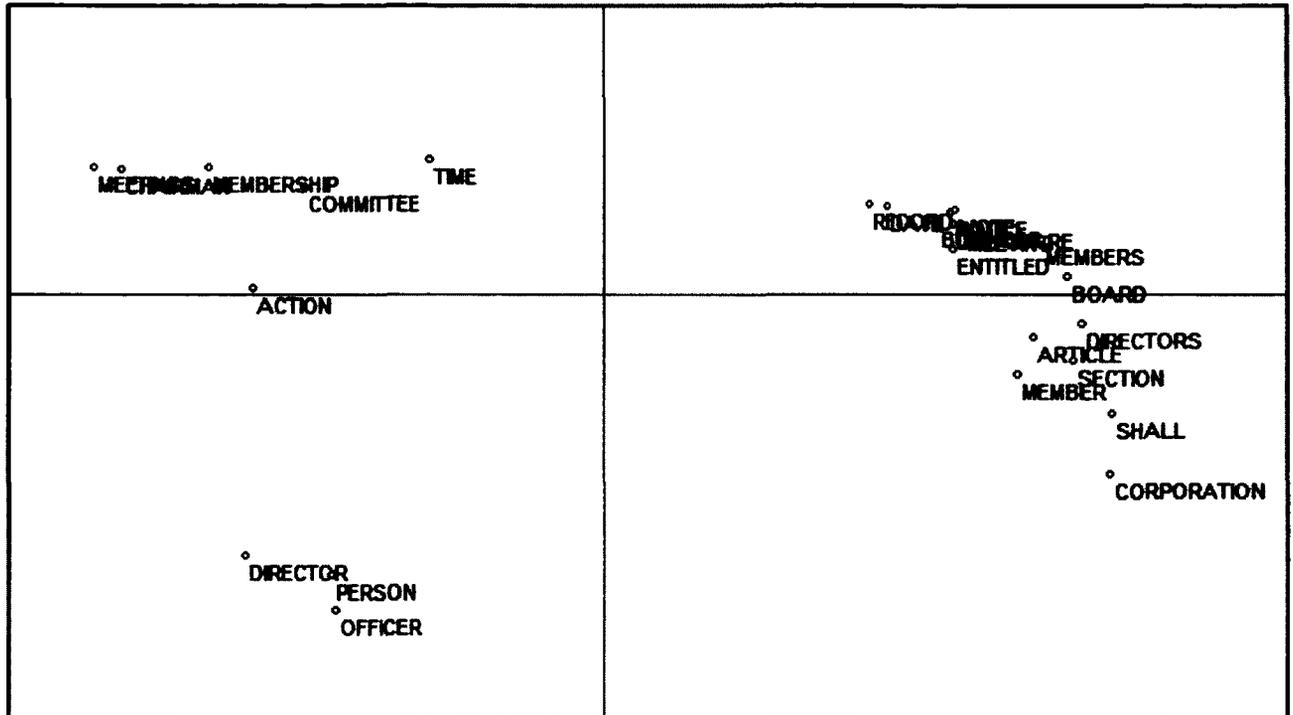


Figure 2. ASF 2D Conceptual Map

ASF Three Dimensional Conceptual Map

Figure 3, the 3D conceptual map, illustrates the relationship between all the concepts. It confirms what has been seen in the other two Figures that the concepts are generally loosely tied together. No one set of concepts have a tight relationship between them.

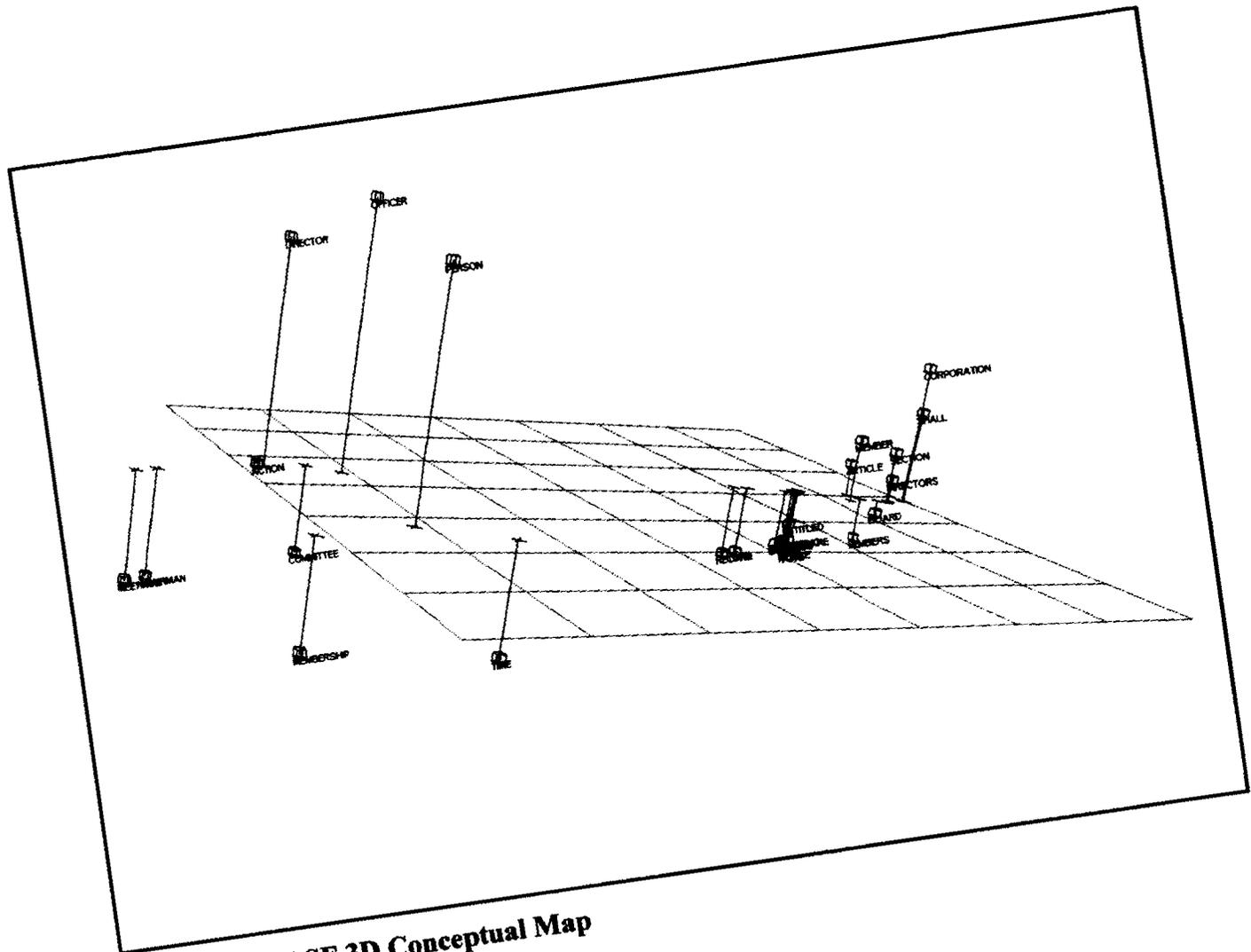


Figure 3. ASF 3D Conceptual Map

4.1.2 The Eclipse Foundation

This section describes the results of the analysis of the data for the Eclipse Foundation.

Eclipse Foundation Hierarchical Cluster Analysis (Dendrogram)

Figure 4 illustrates the dendrogram for the Eclipse Foundation. It shows two strong clusters around the EXECUTIVE DIRECTOR (EXECDIREC) and around ECLIPSE FOUNDATION. From the first cluster we note that the EXECUTIVE DIRECTOR has a strong relationship with COMMITTEES, OFFICERS, ACTIONs and COMMITTEES. The SET and FORTH concepts cluster strongly together because they are used heavily in the bylaws in context such as, “set forth in Section 3.8”. This cluster has a strong relationship with the EXECDIREC.

In the cluster around the COMMITTEE concept, the EXECDIREC has a strong relationship with it because the EXECDIREC is involved in setting up committees. For example, in Section 4.1 it reads, “Each committee shall consist of two (2) or more directors nominated by the Executive Director, including ...”

The second major cluster consists of the relationships of a number of other concepts with the ECLIPSE FOUNDATION. It illustrates a close relationship between the concepts of ECLIPSE FOUNDATION and DIRECTORS and the BOARD of DIRECTORS and the work of the foundation in terms of strategic direction and membership.

Eclipse Foundation Two Dimensional (2D) Conceptual Map

Figure 5 illustrates the 2-D conceptual map, showing the clustered data represented over 4 quadrants. This is basically what is illustrated in the dendrogram. The map can be interpreted as their being a left and a right half, where the left half is dominated by the EXECDIREC and the right half is more of the day-to-day functioning of the foundation.

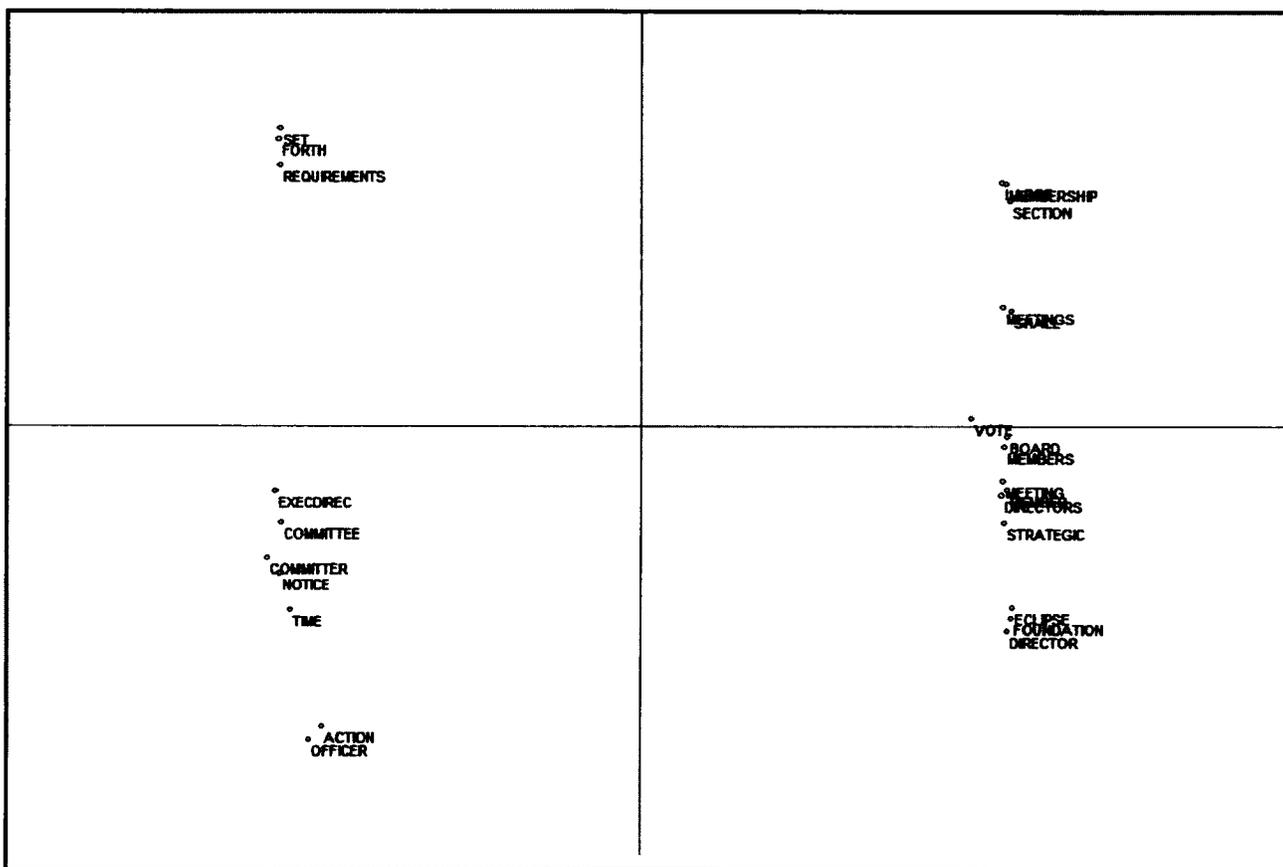


Figure 5. Eclipse Foundation 2D Conceptual Map

Eclipse Foundation Three Dimensional (3D) Conceptual Map

The 3D conceptual map in Figure 6 shows two distinctive clusters, one closely tied to the EXECDIREC and the other to the ECLIPSE FOUNDATION, as with the dendrogram and the 2D conceptual map. The map clearly shows that the strongest relationships in managing the foundation are related to the EXECUTIVE DIRECTOR.

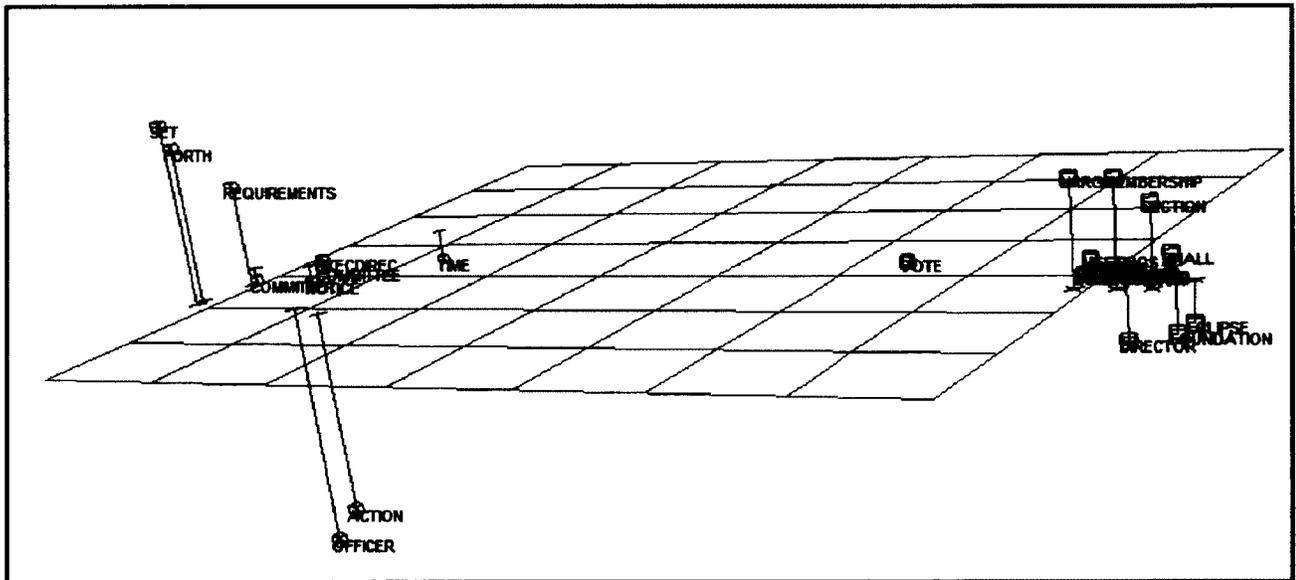


Figure 6. Eclipse Foundation 3D Conceptual Map

4.1.3 The GNOME Foundation

This section describes the results of the analysis for the GNOME Foundation.

GNOME Foundation Hierarchical Cluster Analysis (Dendrogram)

Figure 7 illustrates the dendrogram for the GNOME Foundation. It shows 2 strong clusters around the PRESIDENT and BOARD/DIRECTORS concepts. In the first cluster, we note that the concepts of DIRECTOR, BYLAWS, COMMITTEE and VOTE have a strong relationship with PRESIDENT.

In the second cluster we note that concepts like MEETING, MEMEBERS and CORPORATION have a strong relationship with the concepts of BOARD/DIRECTORS. In Article VIII of the bylaws, we note that the Board acts on the behalf of the CORPORATION (GNOME) in matters such as incurring debt, selecting and removing officers, calling meetings and creating committees. However, in the analysis, the concepts of PRESIDENT and COMMITTEE have a closer relationship than BOARD/DIRECTORS and COMMITTEE. There is no apparent explanation from reading the bylaws other than possibly the BOARD appoints the PRESIDENT and the PRESIDENT acts for the BOARD. This may be an anomaly due to the tool used.

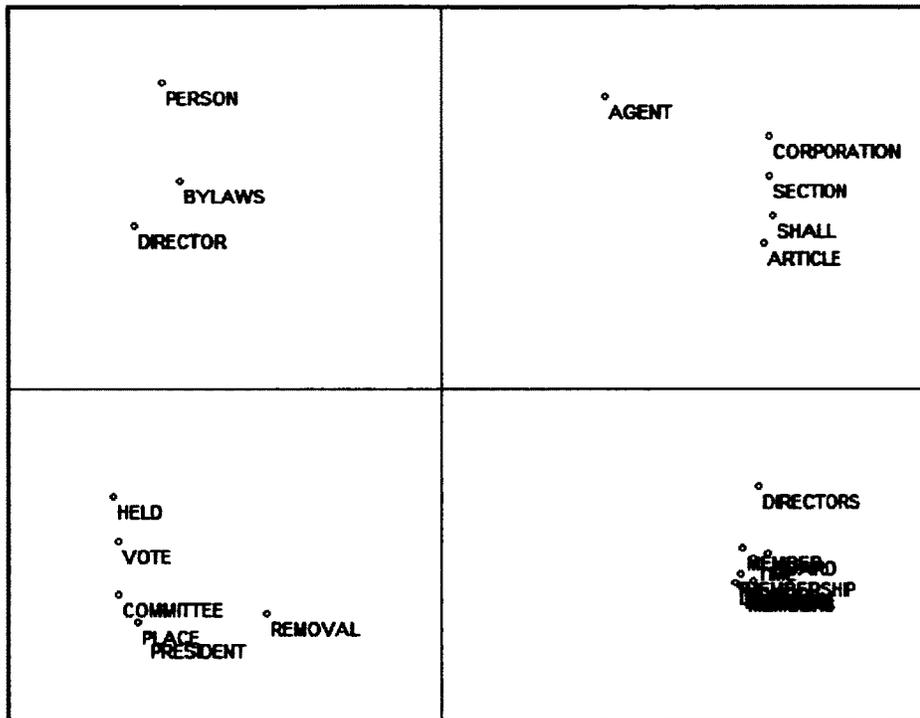


Figure 8. GNOME Foundation 2D Conceptual Map

GNOME Foundation 3D Conceptual Map

Figure 9 illustrates the 3D conceptual map. As with the dendrogram and the 2D conceptual map, the 3D map shows three clusters. No new conclusions about the relationships of the concepts are apparent from the 3D map.

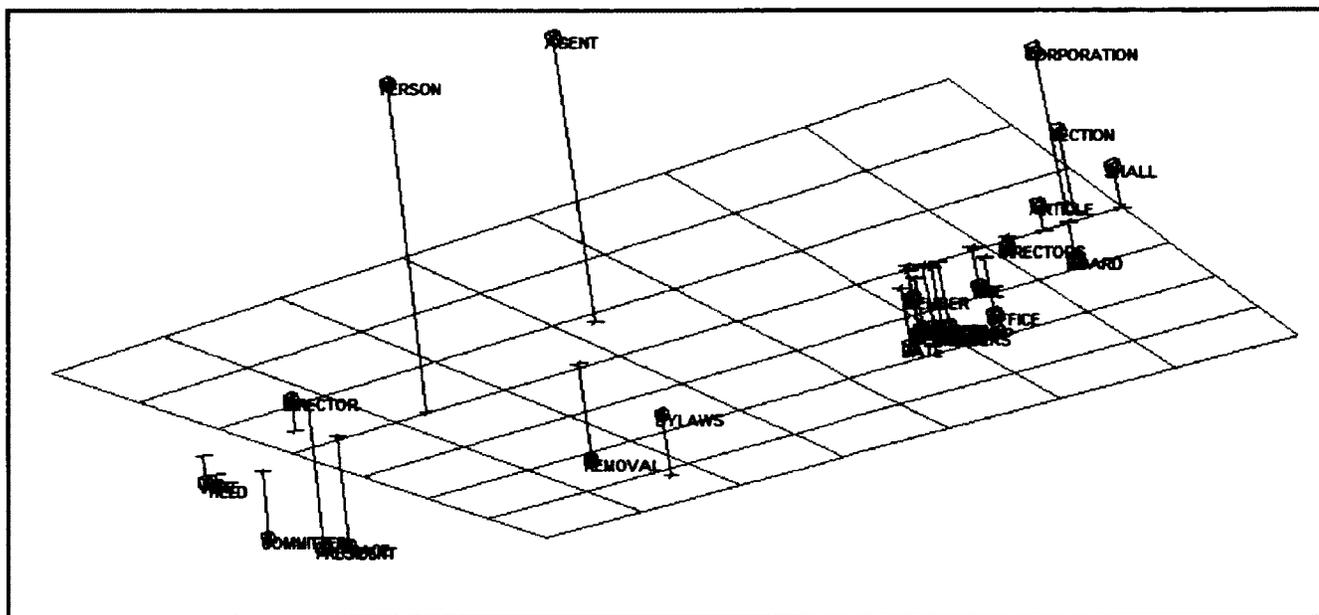


Figure 9. GNOME Foundation 3D Conceptual Map

4.1.4 Plone Foundation

This section describes the results of the analysis for the Plone Foundation.

Plone Foundation Hierarchical Cluster Analysis (Dendrogram)

Figure 10 illustrates the Plone Foundation dendrogram. It shows 2 strong clusters around BOARD/DIRECTORS (BoD) and CHAIRMAN concepts. We note that the BoD concept has a strong relationship with CORPORATION, MEMBERSHIP and MEETINGS concepts. This is attributed to that according to Article 4 of the Plone Bylaws, the BoD manages the CORPORATION's business. This includes calling and conducting meetings, administering the admission and removal of membership, and compensation. In addition the concept of COMMITTEE has a strong relationship with the BoD. As per Article 5 of the bylaws, the BoD establishes committees and appoints members to the committees to "serve during the pleasure of the Board of Directors."

In the bylaws of the Plone Foundation, there is a provision for an Emeritus member. Such a member is one who has left the membership voluntarily or involuntarily. From examining the bylaws, there is no obvious reason for the concepts of CHAIRMAN and EMERITUS to have a close relationship. The same hold true for CHAIRMAN, SECRETARY and OFFICES.

CENTROID METHOD

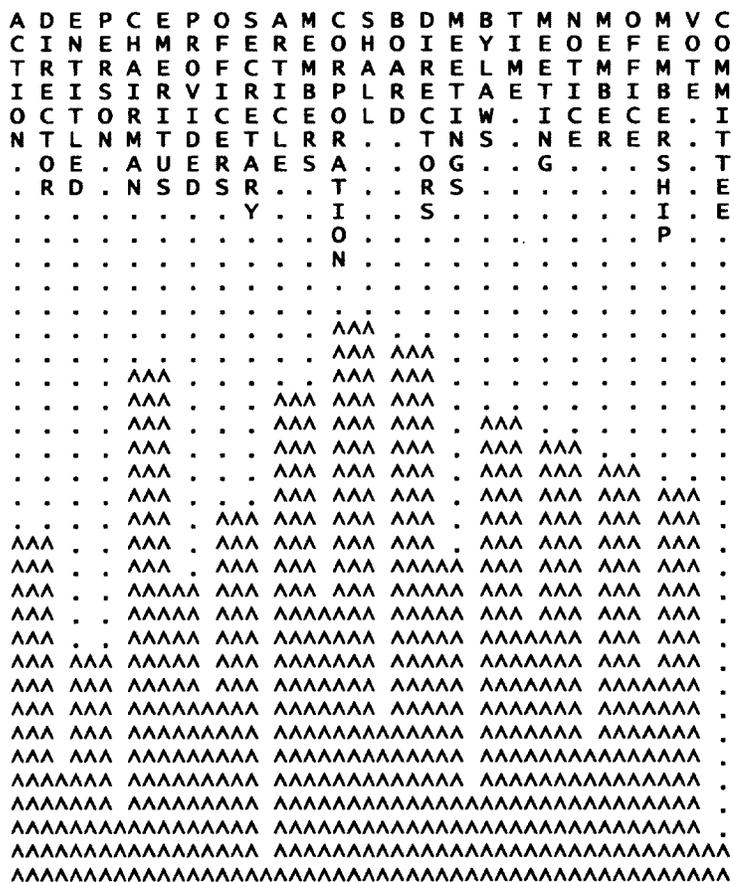


Figure 10. Plone Foundation Dendrogram

Plone Foundation 2D Conceptual Map

Figure 11 illustrates the 2D conceptual map for the Plone Foundation showing the clustered data over 4 quadrants. It is basically showing the same clustering and relationships as the dendrogram.

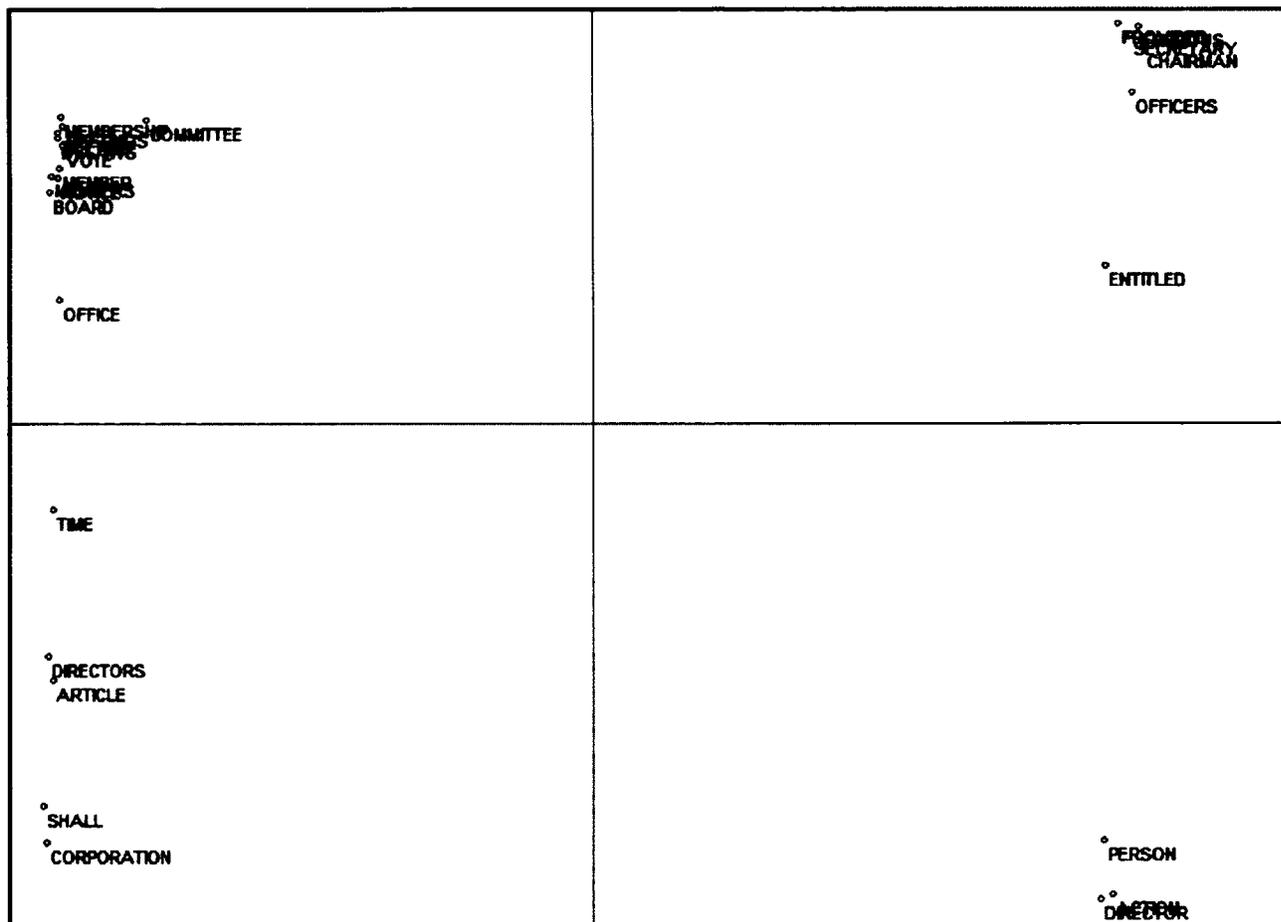


Figure 11 Plone Foundation 2D Conceptual Map

Plone Foundation 3D Conceptual Map

Figure 12 illustrates the 3D conceptual map. As with the dendrogram and the 2D conceptual map, the 3D map shows the same clustering and relationship information as the dendrogram.

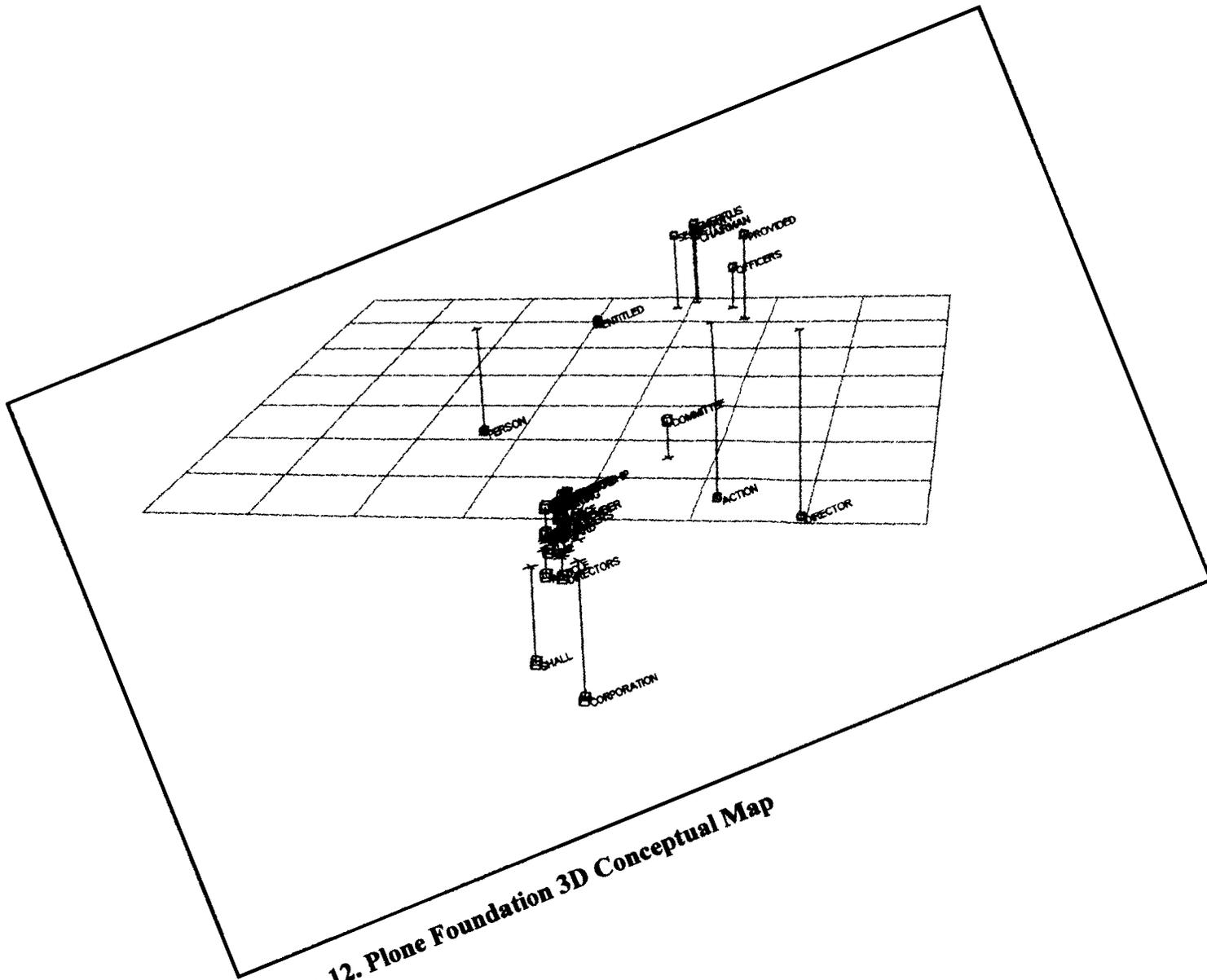


Figure 12. Plone Foundation 3D Conceptual Map

4.1.5 Python Foundation

This section describes the analysis results for the Python Foundation.

Python Foundation Hierarchical Cluster Analysis (Dendrogram)

Figure 13 illustrates the Python Foundation dendrogram. It shows 3 strong clusters around BOARD/DIRECTORS (BoD), MEMBER and CHAIRMAN.

One notes that the BoD concept clusters strongly relationship with CORPORATION and MEMBERS. This reflects Article V of the bylaws, which states, “The business and affairs of the corporation shall be managed by or under the direction of the Board of Directors ...”. In addition, the BoD concept has a strong relationship with the MEMBER and MEETING concepts, the two of which cluster strongly together. Article III of the bylaws sets the rules for holding members’ meetings, clearly stating that the BoD sets the meetings and that the BoD is elected from the MEMBERS. In Article IV, the BoD has responsibility of admitting the initial MEMBERS to the CORPORATION, further strengthening the relationship between the membership and the Board of Directors.

The third strong cluster is around CHAIRMAN. The concept of PRESIDENT clusters with CHAIRMAN likely because in the bylaws, the PRESIDENT can replace the CHAIRMAN of the BoD, if the CHAIRMAN is not elected or cannot be present at meetings.

Python Foundation 2D Conceptual Map

Figure 14 illustrates the 2D conceptual map for the Python Foundation, showing the clustered data over 4 quadrants. Other than the COMMITTEE concept having the closest relationship to the BoD, the information depicted in the 2D conceptual is much the same as the dendrogram.

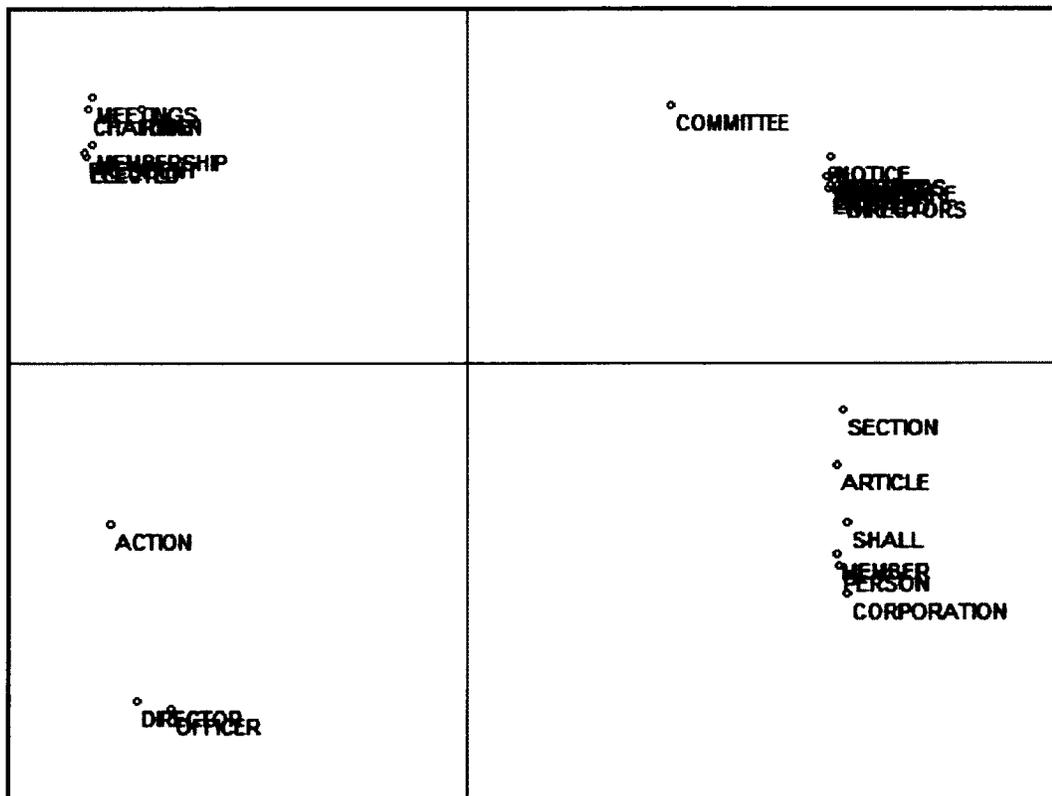


Figure 14. Python Foundation 2D Conceptual Map

Python Foundation 3D Conceptual Map

Figure 15 Illustrates the 3D conceptual map for the Python Foundation. As with the 2D map, the 3D map shows mostly the same as the dendrogram. However, it shows that the COMMITTEE concept is somewhat equidistant between BoD/MEMBER and CHAIRMAN concepts. This is somewhat true also of the dendrogram as there is only one caret difference between CHAIRMAN and BoD/MEMBER. In the 2D case, the relative distance between COMMITTEE and the other two concepts is also somewhat equidistant.

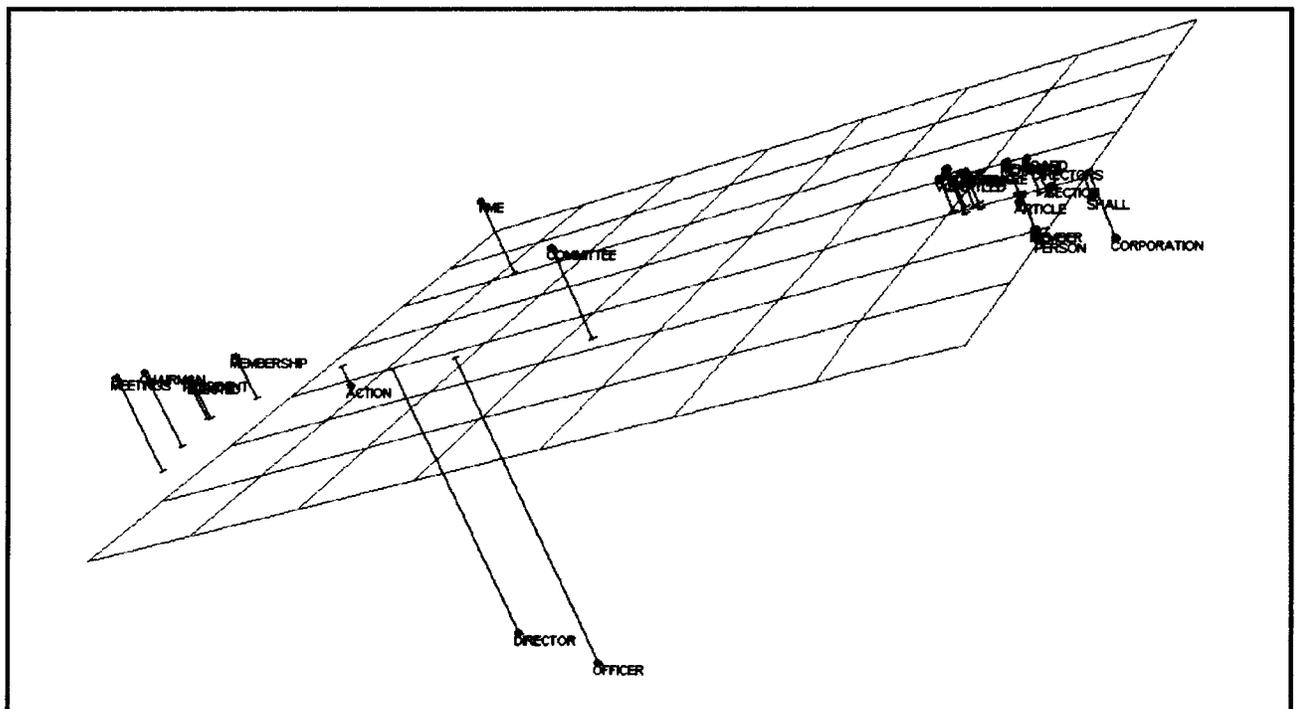


Figure 15. Python Foundation 3D Conceptual Map

4.1.6 Software in the Public Interest Foundation

This section describes the results of the data analysis for the Software in the Public Interest (SPI) Foundation.

SPI Foundation Hierarchical Cluster Analysis (Dendrogram)

Figure 16 illustrates the dendrogram for the SPI Foundation. It shows two strong clusters around the concepts of BOARD/DIRECTORS (BoD) and MEMBER. The clustering consists of smaller clusters around BoD and MEMBER. This is likely due to the SPI Foundation bylaws are relatively short, 8 pages, compared to the other foundations where they are anywhere between 15-25 pages.

Article Three, at the beginning of the bylaws, states that the members are responsible for overseeing the board, offices and committees. The BoD and MEMEBERS illustrate this in the BoD cluster and at a higher lever, with the MEMBER concept.

In Article Seven of the bylaws, we learn that the "... business of the organization shall be managed by the Board of Directors ...". In the dendrogram, we see that the BUSINESS concept has a closer relationship with PRESIDENT. This implies that the president has more to do with the managing of the business than the BoD. A closer examination of the bylaws, specifically in the second paragraph of Article Eight, we note that the president

presides at the membership meetings, is the “chairperson of the board”, “appoint[s] all committees, temporary or permanent”, along with other duties.

In the case of the MEMBER concept, we note it has a close relationship with PRESIDENT, OFFICER, SECRETARY and COMMITTEES. This is consistent with the bylaws in that they say the president, secretary and committees are to be formed from members of the Foundation.

One interesting point is that the CONTRIBUTING concept has a close relationship with COMMITTEES. This suggests that only contributing members can be in committees, but this does not show up in any of the bylaw articles.

CENTROID METHOD

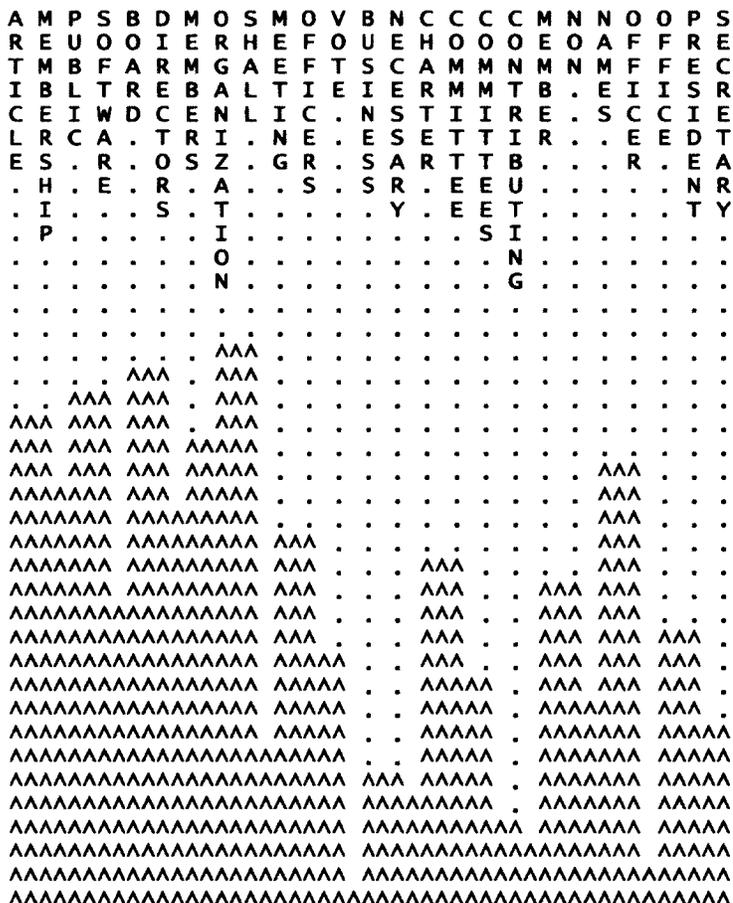


Figure 16. SPI Foundation Dendrogram

SPI Foundation Two Dimensional (2D) Conceptual Map

Figure 17 illustrates the 2D conceptual map for the SPI Foundation bylaws showing the clustered data over 4 quadrants. What we note from the 2D map is that the concept of SECRETARY does not have as close a relationship with PRESIDENT as in the dendrogram. In the bylaws, the secretary is responsible for “presenting to the

membership”, record keeping, filing certificates, “serving all notices to members of the organization”, among others. These are similar to actions associated with the president, hence they cluster in the dendrogram. However, the 2D conceptual map filters this in that it tells us the SECRETARY has a closer relationship with the BoD as it is relatively closer to the concept of DIRECTORS than PRESIDENT.

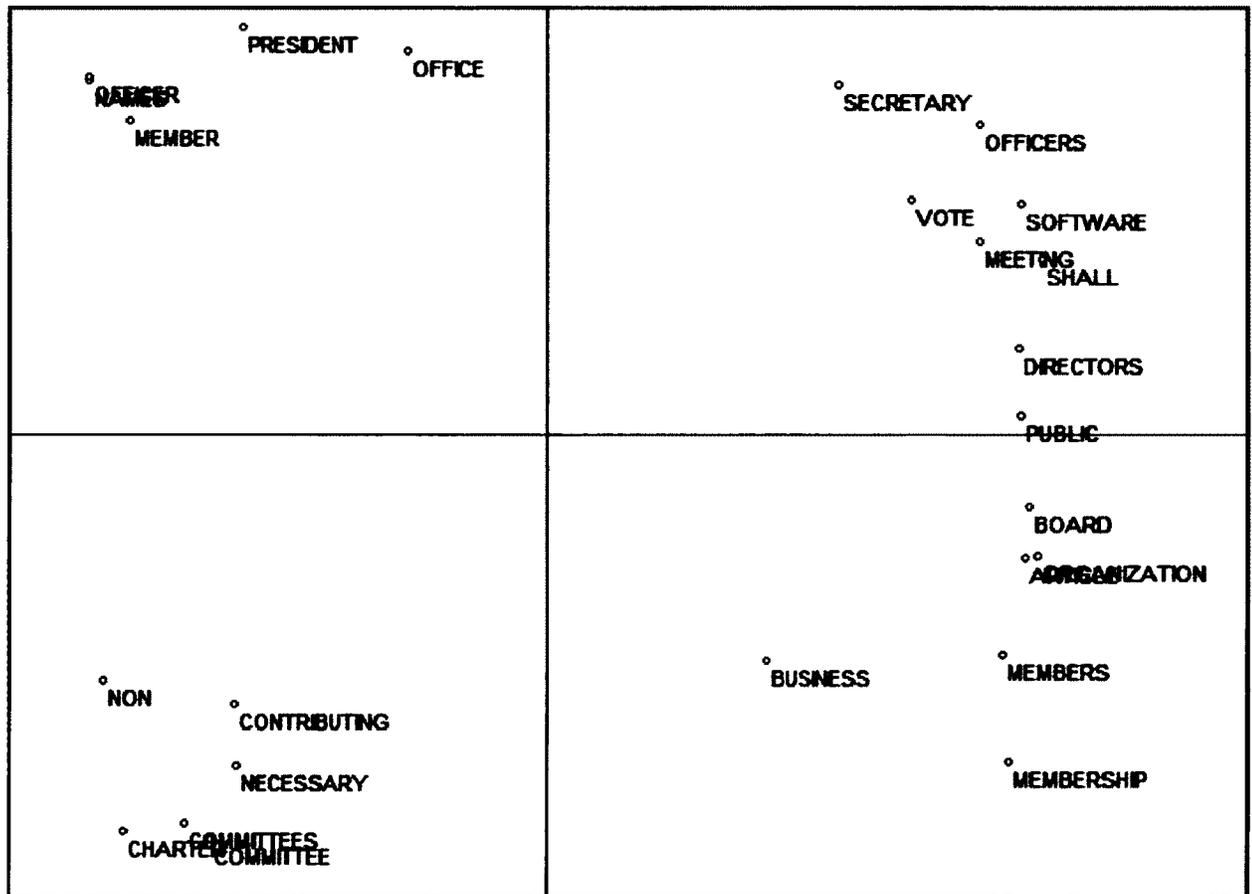


Figure 17. SPI Foundation 2D Conceptual Map

SPI Foundation Three Dimensional (3D) Conceptual Map

Figure 18 illustrates the 3D conceptual map for the SPI Foundation bylaws. Clustering occurs around the concepts of BoD (as defined in the previous section), PRESIDENT and MEMBER. The BoD concept is consistent with both the 2D conceptual map and the dendrogram. However, we note that the concepts of BUSINESS, COMMITTEE, and SECRETARY cluster around PRESIDENT. As noted for the dendrogram analysis, according to Article Eight, the president manages the business of the Foundation. Again the secretary has duties similar to the president, hence the SECRETARY concept would cluster closely with the PRESIDENT concept.

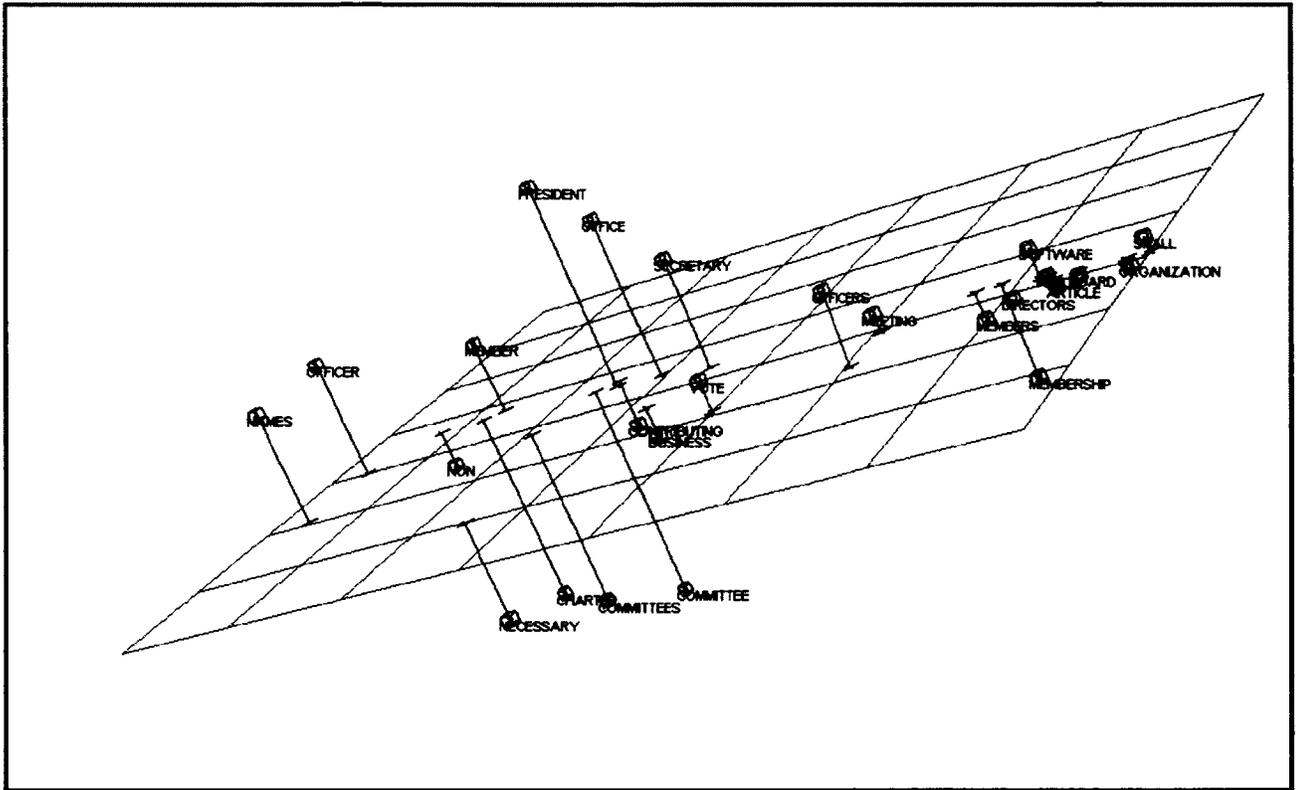


Figure 18. SPI Foundation 3D Conceptual Map

4.2 Summary of results

This section provides an overall summary of the results.

In the case of the Apache Software Foundation, all the concepts were loosely connected but most of them cluster, albeit loosely, around the BOARD OF DIRECTORS concept. As a board consists of members from various members, this is in line with Apache's mission to support collaborative, consensus-based organisation. The implication of this is that in the bylaws, Board of Directors acts on the actions and policies as outlined in the bylaws.

The Eclipse Foundation has stronger clustering of concepts around the EXECUTIVE DIRECTOR concept. This is clearly demonstrated in the 3D conceptual map, Figure 6. This implies that the EXECUTIVE DIRECTOR figures prominently in the actions and policies outlined in the bylaws. The statement of purpose of the Eclipse Foundation say the foundation is to be a steward of the community. It is not evident how a foundation where the actions of the Executive Director figure prominently in the bylaws can act as a steward of the community.

The analysis of GNOME foundation bylaws shows strong clustering around the BOARD OF DIRECTORS in all three visual plots. However, there are clusters that show relationships with the PRESIDENT and MEMBER concepts. This indicates that according to the bylaws, although the Board of Directors has prominence in the management of the foundation, the President and Members play a role in the decision-making.

Plone, like GNOME, shows similar clustering around the BOARD OF DIRECTORS concept in all three visual plots. The implication is that the Board of Directors has prominence in the management of foundation.

The visual plots for the Python Foundation, the BOARD OF DIRECTORS concept clusters strongly with MEMBER and CORPORATION. This again reflects the bylaws in that the Board of Directors manages the “business and affairs of the corporation.”

All three visual plots for the SPI Foundation’s bylaws show strong clustering around the BOARD OF DIRECTORS concept. As with the other foundations studied, this can be related to the bylaws that contain the explicit statement, “ ... [the] business of the organization shall be managed by the Board of Directors.”

5 DISCUSSION OF RESULTS

This chapter is organised into four sections. Section 5.1 provides a categorisation of the OSSF based on content analysis of their bylaws. Section 5.2 answers the research question. Section 5.3 links the results with the literature. Section 5.4 discusses the anomalies. Finally, Section 5.5 discusses the insights from using the tool and for those interested in participating in open source software foundations.

5.1 Categorisation of foundations using bylaws

Figure 19 illustrates the categorisation of the foundations based on their bylaws, as interpreted in Chapter 4. From the analysis we have learnt that the power in each of the foundations lies in different groups. In the case of the Apache Foundation, the Board of Directors has the majority of the power with a skew towards the Members and away from the Chairman/President. The same holds true for the Plone Foundation. However, the power is skewed a little farther away from the Chairman/President than for the Apache Foundation. In the case of the Python Foundation, the power is squarely on the Board of Directors-Members axis. In the case of the Eclipse Foundation, we note that power lies clearly with the Chairman/President (in this case the Executive Director). In the case of the GNOME and SPI Foundations, the power is more centered with a skew towards the Board of Directors and Members.

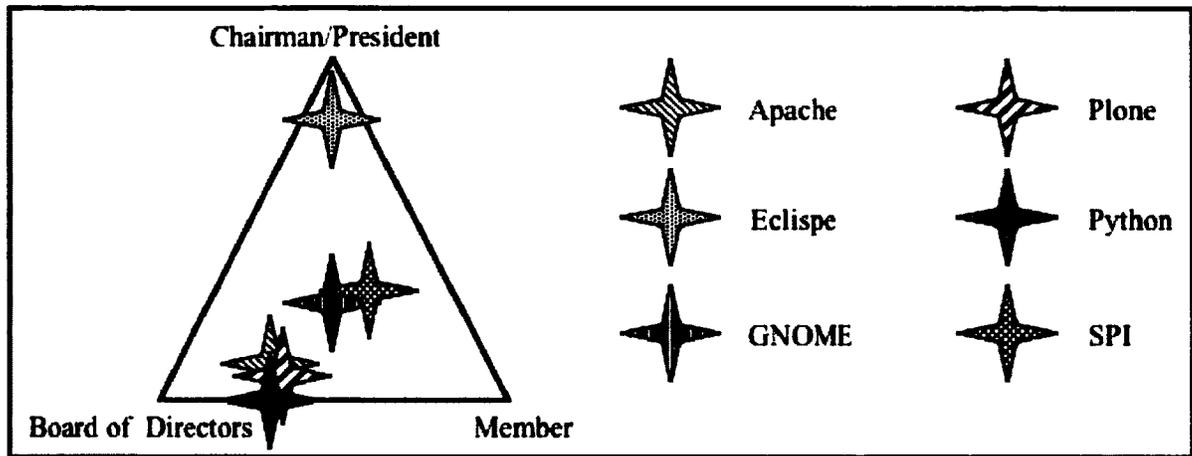


Figure 19. Categorisation of foundations

5.2 Answering the research question

This research studied six open source software foundations that have at least one open source software project and are operating in the same legal jurisdiction. The objective was to study if the management of foundation is centred according to the policies set out in the bylaws. Using computer aided text analysis, it was determined that the actions of an open source software foundation are centered on one of three groups: Members, Chairman/President/Executive Director, and Board of Directors. However, this research did not study why power lies within different groups in the foundations.

5.3 Categorisation of OSSF

Table 7 contains the characterisation of the OSSFs studied by mission statement orientation and where the power is centred according to the bylaws. Effectively, what the characterisation tells us is that the Board of Directors in the Apache Foundation have responsibility towards the community according to the mission statement. In the case of the Eclipse Foundation, the Executive Director has a responsibility towards the community. In the case of the GNOME, Plone and SPI foundations, the Board of Directors have responsibility towards the product. Only in the Python Foundation, does the Board of Directors have responsibility towards the community and the product.

| OSSF | Mission Statement Orientation | Power Centre |
|-------------|--------------------------------------|---------------------|
| Apache | Community | Board of Directors |
| Eclipse | Community | Executive Director |
| GNOME | Product | Board of Directors |
| Plone | Product | Board of Directors |
| Python | Community & Product | Board of Directors |
| SPI | Product | Board of Directors |

Table 7. Mission statements characterisation based power centre

5.4 Links to the literature

O'Mahoney (2007) identified independence of any one sponsor as one of the five principles for community managed open source software communities. This research,

through the analysis of bylaws notes that there are OSSFs where the power is centered with the Chairman/President/Executive Director, are potentially not independent of any one sponsor.

Mitchell et al. (1997) argue that the Freeman (1984) definition stakeholders is ambiguous and not really suitable to identify stakeholders. To better identify the stakeholders, they develop the concept of stakeholder salience described by three attributes: power, legitimacy and urgency. This research examines the bylaws to determine where the power lies based on the bylaws.

Frooman (1999) concluded that stakeholders' influencing strategy is dependent on the resource relationship between the stakeholder and the firm. This research looked at how management of the firm is influenced by the bylaws and concludes that the firm can be influenced depending on how the bylaws centre the power.

5.5 Anomalies in the research

In the analysis of the GNOME Foundation bylaws, we noted that the concepts of PRESIDENT and COMMITTEE had a closer relationship than BOARD/DIRECTORS and COMMITTEE. There is no evidence in the bylaws that this is the case. This may be an effect of using artificial neural network analysis. We have already noted that artificial neural network analysis may not work properly with formal documents (Evans et al, 2010) and this maybe another manifestation of the same issue.

5.6 Challenges and insights of using computer aided text analysis

The tool used is known as CATPAC. CATPAC is neural network software used to establish patterns within written text and draw ideas out of the text (Woelfel & Stoyanoff, 1998). There are several issues with the tool that need to be taken into consideration. This tool was used to eliminate researcher bias and coding errors. This meant that the first few runs of the tool used the straight text input that was converted from the DOC, HTML, and PDF file types downloaded from the Internet. The problems noted included

- i. Text in headers and footers repeated throughout the document
- ii. Words that must be interpreted together were not, e.g. Executive Director
- iii. Words with interpretations that are context sensitive
- iv. Characters not readable to humans

The above cases caused graphical results that were difficult to interpret and in some cases the clustering did not make sense with what was written in the bylaws. To improve on interpreting the results and maintain the original intent of no researcher bias, this researcher performed minimal editing of the text.

Case (i) caused high counts around such words as *bylaws* and the foundation name, skewing the results by causing clustering around the word *bylaws* or the foundation name. To eliminate the problem, the text in the headers and footers was removed before converting the files to pure text.

Case (ii) was particular to the Eclipse Foundation where the words *Executive Director* and *Director* were used throughout the document and meant different concepts. The software did not differentiate the two. To eliminate the issue, the words *Executive Director* were combined to form *ExecDirec*, a new concept and to differentiate it from the use of the word *Director*. This edit was undertaken because it required no interpretation from the researcher because the words *Executive* and *Director* were always next to each other, hence a simple FIND and REPLACE operation with a word processor made the change. In the case of the term *Board of Directors*, it was found that the terms *Board* and *Director* would cluster very closely hence editing was not necessary. No similar issues were found in the other bylaws examined.

Another variation of this case is the concepts of *contributing member* and *non-contributing member*. This issue is really two issues disguised as one. The disguised issue is the words *non* and *contributing*. The software ignores punctuation and treats the concept as two separate concepts. There were no substitutions made for this variation as it involves interpretation by the researcher.

Case (iii) is applicable to all the bylaws. The most prominent example of this was *Chairman*. The concept was used in conjunction with *Chairman of the Board* and *Chairman of a committee*. This researcher chose not to make changes in the text as per Case (ii) because it required some interpretation on the part of the researcher and this would violate the original intent of not injecting researcher bias into the work.

A variation of this is the use of the concepts of *member* and *members* in the bylaws. Careful interpretation by the researcher would be required to eliminate this issue.

Case (iv) took a fair amount of analysis to understand what had transpired. The issue arose following the analysis of the Plone and Python bylaws. They are essentially written in the same fashion, with practically the same words. Hence one would have expected very similar data clustering and conceptual maps. However this was not the case. By examining the bylaw text and the same methods to analyse the text of the United Nations Convention on Human Rights (Evans, 2010) in different languages, the possibility existed that non-printing characters not intended to be in the file, were in the file. The ANN software would have seen these characters as valid text and included them in the analysis. The bylaws were regenerated and the issue was corrected.

5.7 Insights for stakeholders

Mission statements only express the intended purpose of the OSSF and do not predict how it will operate. The structure of the bylaws is a better indicator of how the OSSF behaves than its mission statement.

This research examined six foundations operating successfully for at least two years, in a common jurisdiction, operating under the same laws. Whether bylaws centre the power

in the OSSF with the Chairman/President/Executive Director, Members or the Board of Directors, the OSSF can be successful.

As seen by the weak clustering around the concepts studied, power is not absolute. While a Chairman or Executive Director may have the majority of the power (strong clustering), Members and the Board of Directors can have limited power (weak clustering).

6 CONCLUSIONS, LIMITATIONS AND FUTURE RESEARCH

This chapter is organised into three sections. Section 6.1 provides the conclusions of the research. Section 6.2 describes the limitations of the research are listed and Section 6.3 provides suggestions for future research.

6.1 Conclusions

The research draws three main conclusions.

The first conclusion is although open source software foundations state a clear purpose for their creation; analysis of the bylaws for where the power is centred does not represent the purpose of the OSSF, in a direct manner.

The second conclusion is that the majority of the literature on OSSF governance is on the mechanisms and processes used to manage open source software foundations. More work is needed on how the bylaws impact the management of the OSSFs.

The third conclusion is that artificial neural network analysis of raw bylaws provides unbiased insights on the power structure of OSSFs. However, pre-processing the bylaws with a standardised dictionary of terms, would likely improve the results.

6.2 Research Limitations

This researcher found at least four limitations with the research. First, the research studied where the power lies, but not why it lies within different groups in the foundations studied.

Second the final interpretation of the results is still subjective and influenced by the researcher's knowledge and experience. This is because interpretations of the dendrograms and conceptual plots rely on the reasons the researcher deduces from knowledge of the bylaw contents and the roles of the Board of Directors, President/Chairman/Executive Director and the Members.

Third, the bylaws of six open source software foundations were sampled. Eisenhardt (1989) recommends sample sizes of between 4 and 10. According to the Eisenhardt recommendation a sample size of six is sufficient. It is possible that the sample is small because there is only one foundation that is President/Chairman/Executive Director centred, and five that are centred somewhere between the Board of Directors and Members. According to Eisenhardt's recommendation, this may not be enough to fully prove-in the use of ANN analysis in this domain. However, this researcher concluded that the sample size is sufficient to demonstrate that worthwhile information can be obtained from bylaws using ANN analysis.

Lastly, there is evidence that ANN analysis does not work very well with documents written in formal language (Evans et al, 2010).

6.3 Suggestions for Future Research

There are three directions for future research. Future research in understanding why the power lies within different groups can be undertaken.

This research let the tool pick out the concepts that clustered. Instead, study how related concepts cluster around the concepts of CHAIRMAN/PRESIDENT/EXECUTIVE DIRECTOR, BOARD OF DIRECTORS and MEMBERS, by forcing the tool to look at the clustering around these concepts only. Some pre-coding would be involved in perform this research.

A second suggestion is to pre-process the bylaws in much the same way Samkin & Schnieder (2008) did, standardising the language used and investigate how they cluster.

The third suggestion is to analyse the bylaws through the application of stakeholder theory. The bylaws can be analysed to identify who of the members, board of directors or executive directors are the salient stakeholders and which of the power, legitimacy and urgency attributes they manifest (Mitchell et al, 1997). In turn, this will give a strong indication of where the power lies.

References

Allen, S.D. 2005. Using perceptual maps to communicate concepts of sustainable forest management. *The Forestry Chronicle*, 81(3): 381-386

Bailetti, A.J. & Callahan, J.R. 1995. Managing consistency between product development and public standards evolution. *Research Policy*, 24: 913-931.

Dahlander, L. & Wallin, M.W. 2006. A man on the inside—Unlocking communities as complementary assets. *Research Policy*, 35: 1243-1259.

de Laat, P.B. 2007. Governance of open source software: state of the art. *Journal of Management & Governance*, 11:165–177

Doerfel, M.L. & Marsh, P.S. 2003. Candidate-issue positioning in the context of presidential debates. *Journal of Applied Communication Research*, 31(3): 212-237.

Eisenhardt, K.M. 1989. Building theories from case study research. *Academy of Management Review*, 14(4): 532-550.

Evans, C., Chen, H., Battleson, B., Wölfel, J.K. & Woelfel, J. 2010. ANN multilingual pattern recognition. Rah Press. <http://www.galileoco.com/literature/Wolfpak10a.pdf>, last accessed June 2012.

Fitzgerald, B. 2006. The transformation of open source software. *MIS Quarterly*, 30(3): 587-598.

Freeman, R.E. 1984. *Strategic management: a stakeholder approach*. Boston: Pitman.

French, J.R.P. and Raven, B. 1959. The bases of social power, in I.G. Asherman & S.V. Asherman (eds.) *The Negotiation Sourcebook* (2nd ed.). Amherst, MA: HRD Press.

Frooman, Jeff. 1999. Stakeholder Influence Strategies. *Academy of Management Review*, 24(2): 191-205.

George, G. & Bock, A.J. 2011. The Business Model in practice and its implications for entrepreneurship research. *Entrepreneurship Theory and Practice*, 35: 83–111

Jin, L., Robey, D., & Boudreau, M.-C. 2007. Beyond development: A research agenda for investigating open source software user communities. *Information Resources Management Journal*, 20(1): 68-80.

Milinkovich, M. 2010. How can a community be considered "open source" if its primary objective is to promote commercialization? *Open Source Business Resource*, January 2010. Available at <http://www.timreview.ca/article/320>, last accessed June 2012.

Neuendorf, Kimberly A. 2002. *The content analysis guidebook*. Thousand Oaks, CA: Sage Publications.

Mitchell, R.K., Agle, B.R. & Wood, D.J. 1997. Toward a theory of stakeholder identification and salience: Defining the principle of who and what really counts. *Academy of Management Review*, 22(4): 853-886.

OED – Oxford English Dictionary, Second edition, 1989; online version June 2012. <http://www.oed.com/view/Entry/49943>; accessed 17 July 2012. First published in A Supplement to the OED I, 1972.

O'Mahony, S. 2007. The governance of open source initiatives: what does it mean to be community managed? *Journal of Management & Governance*, 11: 139-150.

Pfeffer, Jeffery. 1981. *Power in Organisations*. Marchfield, MA: Pittman Publishing Inc.

PSF Mission Statement, Python Software Foundation, <http://python.org/psf/mission/>, last accessed July 2012.

Riehle, D. 2010. Economic case for open source foundations. *IEEE Computer*, January 2010, 43(1): 93-97.

Ryan, Chris. 1998. Saltwater Crocodiles as Tourist Attractions. *Journal of Sustainable Tourism*, 6(4): 314-327.

Samkin, G. & Schneider, A. 2008. Comparing a New Zealand and an International Accounting Standard. *Conference Papers of Accounting and Finance Association of Australia and New Zealand*. <http://www.afaanz.org/openconf/2008/openconf.php> Last accessed June 2012.

Stemler, Steve (2001). An overview of content analysis. *Practical Assessment, Research & Evaluation*, 7(17). Available at <http://PAREonline.net/getvn.asp?v=7&n=17>. Last accessed 20 July 2012.

Taft, D.K. 2009. Is Eclipse an Open-Source Community or Trade Association? *eWeek.com*. <http://www.eweek.com/c/a/Linux-and-Open-Source/Is-Eclipse-an-Open-Source-Community-or-Trade-Association-895397/>, last accessed June 2012.

Teece. D.J. 1986. Profiting from technological innovation. *Research Policy*, 15(6): 286-305.

Teece, D.J. 2006. Reflections on “Profiting from Innovation”. *Research Policy*, 35: 1131-1146.

Woelfel, J. & Stoyanoff, N.J. 1998. *CATPAC: A neural network for qualitative analysis of text*. Amherst, NY: RAH Press.

Wölfel, J.K. 1998. *CATPAC II™ Users’s Guide: ver 2.0*. Amherst, NY: RAH Press.

Xie, Z. 2008. *Open source software foundation: company involvement, governance, and effectiveness*. Masters of Applied Science Thesis. Carleton University, Ottawa, Canada.

Appendix A. CATPAC II and ThoughtView Software

(Note parts of this appendix are adapted from the CATPAC II™ User's Manual.)

To perform the analysis, two software applications are used. The first one is CATPAC II, software that performs qualitative text analysis using artificial neural networks and generates the clustering. The second application used is ThoughtView, used to generate two-dimensional and three-dimensional conceptual maps.

CATPAC is a software implementation of a self-organising artificial neural network (ANN), optimised for analysing text (Wölfel, 1998.) It can identify the most important words in a given text and determine patterns of similarity based on how the words are used. It does this by assigning a neuron to each major word in the text. It then runs a scanning window through the text. The neuron representing a word becomes active when that word appears in the window, and remains active as long as the word remains in the window. Up to n words can be in the window at once, where n is a parameter set by the user. As in the human brain, the connections between neurons that are simultaneously active are strengthened following the law of classical conditioning. The pattern of weights or connections among neurons forms a representation within CATPAC of the associations among the words in the text. This pattern of weights represents complete information about the similarities among all the words in the text. Technically, the pattern of connections among neurons is a complete paired comparison similarities matrix, and so lends itself to the most powerful and sophisticated of statistical analyses. The analysis

methods available through CATPAC are various hierarchical cluster analysis methods and multidimensional scaling of words that provides information regarding the relationships amongst the words.

The visual outputs of CATPAC and ThoughtView are *dendrograms*, *two-dimensional (2D) conceptual maps* and *three-dimensional (3D) conceptual maps*. Dendrogram or hierarchical clustering diagrams describe the relationships between the most commonly occurring concepts in the text undergoing analysis. Across the top of the dendrogram are the unique words in the text. The shading (^) underneath demonstrates the words that cluster together. The height of the shading gives an indication of the strength of the clustering. A label that best represent the cluster's theme can be assigned to the cluster, by the researchers based on the researcher's opinion.

ThoughtView creates visual representations of the data by extracting the first three dimensions of the neural network built by CATPAC. In both the 2D and 3D graphs, words that cluster together represent emergent meaning in the text (Doerfel & Marsh, 2003). The 3D visuals are best viewed on a computer monitor where they can be expanded and rotated in all 3 axes to get a better understanding of the clustering.

Hierarchical clustering analysis and multidimensional scaling represent two different perspectives of viewing the data. This manifests itself in that words that cluster together using hierarchical clustering may not cluster together using multidimensional scaling.

This is because hierarchical cluster analysis clusters words based on their co-occurrence and multidimensional scaling algorithms “represent the continuous and multilayered nature” of the neural network through its underlying dimensions (Doerfel & March, 2003). On the other hand, multidimensional scaling algorithms represent the continuous and multilayered

CATPAC II and ThoughtView are available from www.galileo.com for a fee. This research uses the default settings for CATPAC. The *centroid clustering method* was used to eliminate outlier data.

Appendix B. OSSF Mission Statements Provenance

Not all OSSFs had formal mission statements on their websites. However, those that did not have a formal mission statement had available a statement of purpose that was useful in the analysis. This statement of purpose was used as the mission statement.

Apache Software Foundation

As of this writing, the Apache Software Foundation has no mission statement published on its site. However, they do have a FAQ why it was formed. A formal mission statement was located in a conference presentation by one of its founders, W.G. Stoddard at the *5th Open Source China Open Source World Summit*, 2010 Beijing. It succinctly echoes the FAQ. This mission statement was captured in the Table in Chapter 5.

URL for FAQ: <http://www.apache.org/foundation/faq.html#what>

URL for mission statement:

<http://www.slideshare.net/wgstoddard/open-source-at-the-apache-software-foundation>

Eclipse Foundation

No formal mission statement could be located on the Eclipse Foundation website. Indications are they favour a mission statement on a per project basis. An About Us FAQ was located, on their website which describes the reasons for creating the Foundation.

URL: <http://www.eclipse.org/org/>

GNOME Foundation

The GNOME Foundation has a formal mission statement in their online charter. This was used in the table.

URL: <https://live.gnome.org/Foundation/Charter>

Plone Foundation

Plone has a mission statement on their website, used in this research.

URL: <http://plone.org/foundation/>

Python Foundation

The Python has a brief mission statement on their website, used in this research.

URL: <http://www.python.org/psf/mission/>

SPI Foundation

SPI has no formal mission statement on their website, but their Certificate of Incorporation has a statement which amounts to a mission statement. This statement was used in this research.

URL: <http://www.spi-inc.org/corporate/certificate-of-incorporation/>