An Empirical Examination of Open Standards Development

Rajiv Shah
Department of Communication
University of Illinois at Chicago
rshah@a5.com

Jay P. Kesan
College of Law & Dept. of Electrical & Computer Engineering
University of Illinois at Urbana-Champaign
kesan@law.uiuc.edu

ABSTRACT

This project uses empirical data to provide insights into the impact of open standards. This work moves beyond the existing literature by considering a large number of open standards, instead of handpicked case studies. The results of this research will be timely, as governments are advocating and sometimes mandating the use of open standards. We found inequalities in the impact of open standards that suggest a power law relationship, found that the duration of the development process does not affect the impact of a standard, and found the length of a standard (number of words), which reflects the technical complexity of a standard, affects the impact of a standard.

I. INTRODUCTION

One of the most important issues for information infrastructures is standardization. Standards allow for interoperability and interconnection. They provide "the digital equivalent of a common gauge for railroad tracks" [13]. Much of the success of the information age can be attributed to the use of standards, whether it is for transporting information (TCP/IP, FTP) or representing information (HTML, XML, JPEG). When everyone uses the same standards, there are powerful network effects that increase the utility of a technology for everyone [5].

With the growth and power of standards, there has been a noticeable shift in how standards are produced and used. In the past, standards were created through two means. The first is de facto standards created by an industry. These proprietary standards are not publicly distributed and often require a licensing fee for others to use it. A simple example of a "de facto" standard is the Microsoft Word document format. The alternative method is known as "de jure" standards, which are developed by formal standard-setting organizations, such as the International Organization for Standardization (ISO).

This simple dichotomy does not capture how standards are developed nowadays. Firms, individuals, and government have all recognized the value of standards and developed new organizational forms to collectively develop standards. These forms are often categorized as consortia. Within consortia, there are significant variations in the rules setup to produce standards [23]. However, the standards cannot be categorized simply as de facto or de jure. Instead, there are a variety of new approaches to creating and distributing standards. The most prominent is known as open standards. Although the definition of open standards is quite grey [10, 24], we define open standards based on three criteria [4]. First, the standard is publicly available to everyone at a minimal cost. Second, no entity controls the standard, or the standard is licensed on "reasonable and nondiscriminatory terms." Third, the development process for creating the standard involves public participation. However, open standards should not be confused with "open source." Open source is a development model for software based on the public availability of the source code. Open source developers often work together and build upon each other's work. While open source software typically relies upon and uses open standards, they are two different concepts.

The advantage of open standards is they promote interoperability. Interoperability ensures that products from different vendors will work together. This occurs because open standards lower barriers to entry for developers. This prevents lock-in among users by allowing for competition among multiple venders with compatible technologies. Open standards are often considered a form of a quasi- public good, because of the benefits they confer not only on the immediate participants, but also for all of society [22]. Examples of open standards include the transmission protocols such as FTP, HTML, which serves as the language for Web pages, and the image format known as JPEG. Many observers feel that the strength and popularity of the Internet is partly the result of a reliance on open standards.

Open standards have direct implications on public policy concerning competition. A number of governments and organizations, for instance in the United States [3] and the United Kingdom [14], have called for government policies that strongly encourage or mandate open standards. Open standards are touted as reducing costs by avoiding lock-in due to competition from multiple vendors. Besides lower costs, open standards are touted as accelerating economic growth, efficiency and innovation [13].

The promise of open standards has created a number of issues for policymakers. As a starting point, policymakers are concerned about the effectiveness and impact of open standards. Do they really have an enormous impact in reducing costs, lock-in effects, creating interoperability, and generating innovation? And from a practitioner's viewpoint, if open standards are useful, what strategies can developers and consortia use to increase their adoption and use. These questions also led to debates among activists, scholars, and policymakers about the role that public policy should play (or not). For some, government spending and research should favor open standards over proprietary solutions. Others take a more nuanced view, while some advocate no government intervention or influence on standards development and procurement.

Our objective is to empirically assess the effects of open standards. By collecting quantitative data on open standards, our research provides descriptive and analytical insights into the impact of open standards. The motivation for this research stems from our view that the high impact and highly touted open standards do not represent the norm for open standards. Only by collecting data on a set of open standards can we prove this and then begin to determine the crucial factors that differentiate successful standards from less successful ones. This paper builds off of our earlier work[17]. The paper is organized with a section that provides further background on open standards, followed by methodology, our results, and the implications of our findings.

II. BACKGROUND ON OPEN STANDARDS

As the role of standards in digital government increases, considerable debate has arisen over the most effective methods for developing and distributing standards [6]. The bulk of the rich standards literature has focused on the de facto standard-setting process, such as the standards wars over Betamax and VHS video standards, [18] or analyzing a firm's standardization strategy [1]. The work on the de jure standard-setting process largely focuses on coordination issues [8] or factors that influence the standards development process, e.g., intellectual property policies [12]. Consequently, despite the generous body of literature on the economics of compatibility standards, there is a lack of scholarship on open standards [6, 7].

Much of the research and discussion surrounding open standards largely rest on anecdotal evidence [11, 25]. While a few empirical studies have touched upon open standards, there has not been any significant effort to assess the impact of open standards. For example, Simcoe has studied the development process for open standards within the Internet Engineering Task Force (IETF), but not their impact once in use [20, 21]. As a result, when most scholars and policymakers think about open standards, they envision the wildly popular open standards, such as HTML or TCP/IP.

Open standards are an example of a "governance characteristic" of software [16]. Utilizing an open standard supports interoperability and limits lock-in, thus allowing for multiple implementations. However, despite its popularity and success in Internet specifications, there are two main reasons why open standards are not used for all standards. The first is that the development of open standards process typically takes time. This process can slow down the development and implementation of new technology, and, as a result, firms may be unable to meet the demands of their customers on time. A second reason is that open standards do not allow any party to control the standard. Firms tend to favor the control of a standard, especially if they hold a dominant position in the market [19].

There are two main problems with the current understanding of open standards. First, there is a need for empirically assessing the impact of open standards. The attributes of open standards stand them apart from de facto standards and de jure standards developed in traditional standard developing organizations. Policymakers need empirical validation that open standards are indeed beneficial. Without such evidence, it would be ill-advised to blindly put into place preferential policies that favor open standards. Second, there is no guidance for how standards organizations and consortia can improve the production and adoption of standards. The lack of guidelines has led to a multiplicity of approaches. There is a need for research that assesses how standards develop and provides core strategies to improve the process.

III. METHODOLOGY

There are readily-available cases that illustrate the advantageous impact of open standards. Part of our motivation is our belief that the high impact and highly touted open standards do not represent the norm for open standards. Only by sampling a cross section of standards can we prove this and then begin to determine the crucial factors that differentiate successful standards from less successful ones. This project focused on data on open standards from the IETF. The IETF is widely considered as the exemplar for open standards consortia. This section discusses our approach with an emphasis on the problems of measuring the impact of open standards.

A. General Approach

Our data collection involved two datasets. The first comes from working groups that are deemed concluded. Working groups are supposed to be short-lived and are concluded when they meet their established goals and objectives [2]. The data set consists of 111 working groups (WG) that were concluded in the Applications Area, General Area, Internet Area, IP Next Generation Area, and the Network Management Area. This data set encompasses 634 standards. The second set of data is focused more narrowly. It consists of 32 standards randomly selected standards between the years of 2000 and 2003. Since the data had been collected in 2006, it allows us to assess the long-term impact of an open standard. These standards are identified in the Appendix.

The variables of interest include information on the document, authors, and development process. The variables for the document include the page and word length of the standard, number of references, number of other IETF standards cited, and the working group within the IETF. A second part focuses on the authors. These variables include the number of authors, their affiliations (educational or commercial), and their experience writing other IETF open standard. The third category of the development process includes variables on the length of the development process from the first draft to publication, the number of drafts, and intellectual property disclosures.

The other set of variables focus on the impact of the standard. As we explain in the next subsection, the variables include several different measures of impact. The first is impact based on popularity is gauged through online search engines, such as Google search results. The second focuses on scholarly impact by examining the results of a Google Scholar search. The third focuses on citations from other IETF RFCs. The final measure is based upon citations from patent database.

B. Difficulties in Measuring Impact

To understand the difficulties of measuring impact, consider two of the standards in the dataset. The first is RFC 3641, titled, "Generic String Encoding Rules (GSER) for ASN.1 Types." The standard defines a set of ASN.1 encoding rules, called the Generic String Encoding Rules or GSER, that produce a human readable character string encoding of ASN.1 values of any given arbitrary ASN.1 type. In short, this standard seeks to standardize a system of encoding rules, with a benefit to end users of automating these rules. The primary domain for this technology is for LDAP or Lightweight Directory Access Protocol, which is a protocol that email and other programs use to look up information from a server, such as a person's email address. A second standard is RFC 3161, Internet X.509 Public Key Infrastructure Time-Stamp Protocol (TSP). This standard describes the format for messages sent to a Time Stamping Authority, which is a common activity, e.g., email messages. This function is needed, and without this standard, there would undoubtedly be another standard for this activity. This standard replaced many proprietary solutions for time stamping, and now, vendors routinely support 3161 so they can gain the benefits of interoperability. As a result, modern operating systems are all 3161 compliant.

One approach to measuring impact is to focus on how widely a standard is adopted. Factors such as market share could provide quantitative representation of the impact or importance of a standard. For RFC 3641, there are two products that have implemented the standard, View500, which is proprietary and commercial (and not marketed in the US), and OpenLDAP, which is free and open source. There are tens of LDAP products, including ones by major vendors such as Microsoft, IBM, Sun, and Novell. There is no data on the market share of any LDAP product. Moreover, if there were market share data, it would be difficult to untangle the gains for this standard versus other standards, new technologies, changes in customer preferences, and a whole host of other factors that impact the adoption of a technology. As a result, the market share data does not offer a very accurate measure of impact.

In most cases, it is very difficult to determine the number of implementations for a given open standard. Since the standard is open, anyone could implement it. Once implemented, it is not obvious that the technology is incorporated, because most technologies do not list all the standards they incorporate or are compliant. With open source, the market share becomes more vague, because just one person with a handful of code could add support to the Linux operating system. Since Linux is freely distributed, the resulting compliant code could be on millions of desktops. However, there is no way to ascertain whether people are actually using or relying on that standard.

RFC 3161 also highlights similar problems with market share, because it is widely used. How can market share or a dollar value be calculated on a standard that is supported by most modern operating systems? Our preliminary research finds that open standards do not directly compete with proprietary solutions and technologies. Most vendors see open standards as underlying technologies. Proprietary products are built upon open standards, with some proprietary extensions. When an open standard is developed (often by vendors who have cooperated in the development process), it typically pushes aside proprietary solutions. Or, the standard becomes irrelevant, because other standards or technologies outpace it. Instead of a competing solution, open standards are better thought of as akin to nuts and bolts and not as critical differentiators that set products apart from their competitors.

The problems with market share led us to consider other impact measures. The first measure was to ask experts in the field to assess the impact of the standard. This was time consuming and difficult. For several standards, we collected data on the assessment of a standard on a scale of 1 to 10. Experts would often give very different assessments, largely because of a lack of agreement on the proper technical domain to assess. For example, for RFC 3641, is the impact focused on ASN.1 for LDAP, LDAP, directory servers, or search technologies? Another approach was to calculate impact based on citation counts. In academia, the amount of citations for an article is often a proxy for the impact of the article. After trying a number of academic databases, we finally settled on citations from the U.S. patent database, other IETF RFCs, and Google Scholar search results. The final impact measure is more akin to popularity. Our goal was to determine which RFCs fostered the greatest interest and discussion. This popularity measurement was gauged through the number of Google hits for each RFC.

IV. RESULTS OF POWER LAW ANALYSIS

The descriptive data on the impacts of a standard found an interesting relationship. First, Figure 1 ranks standards by their impact (the number of Google hits). From this plot, we can see only a few standards have very high Google hits, while a lot of standards have very small hits. Table 1 shows more detail. 80% Google hits come from only 17.5% of existing standards.

% of Google hits	% of standards
20	0.9
50	4.4
80	17.5

Table 1: Percent of Google hits and the Corresponding Percent of Standards

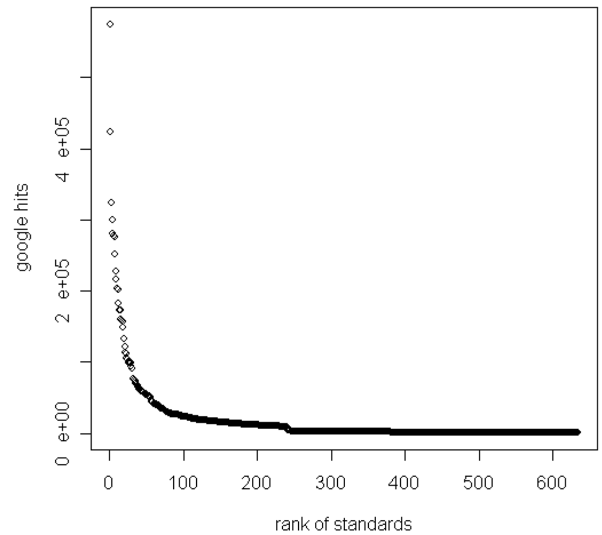

Figure 1. Ranking Standards by Impact

The distribution here is not Gaussian, but follows a power law distribution. According to the definition, the power law relationship exists when:

$$Y = KX^{\alpha}$$

where Y is number of Google hits, X is the rank of the standard, K is a constant, and α is the exponent parameter. To fit this relationship, usually we take logarithm on both sides. And the equation above transforms to:

$$\log Y = \log K + \alpha \log X$$

From this transformation, the power relationship becomes a linear one. A log-log plot does not show us a perfect linear relationship, but that is characteristic of a power law relationship. The analysis finds that the dot trend becomes steeper at the point when $\log X = 5.5$ (X equals 245th) approximately. At this point, the data can be separated into two parts and fit the regressions again as shown in Figure 2. These regressions fit very well, with the upper ranked data and lower ranked data with a slope that equals –0.97 and –2.59 respectively. For the upper ranked data, the –0.97 slope indicates Zipf's Law[1].

[1] Zipf's law states that the size of the r'th largest occurrence of the event is inversely proportional to its rank: $Y \propto R^{-b}$ with ***b*** close to unity.

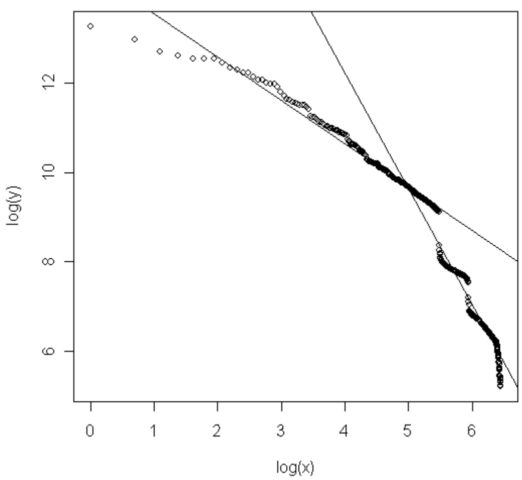

Figure 2. Log-Log Plot of Standards by Impact

We also found a power law relationship among working groups. This part of the analysis used working groups (WG) as the unit of analysis instead of individual standards. A ranking of WG by its popularity (the number of Google hits) is shown in Figure 3. Once again, we see very large Google hits come from only a few WGs, and lots of WGs have very small Google hits

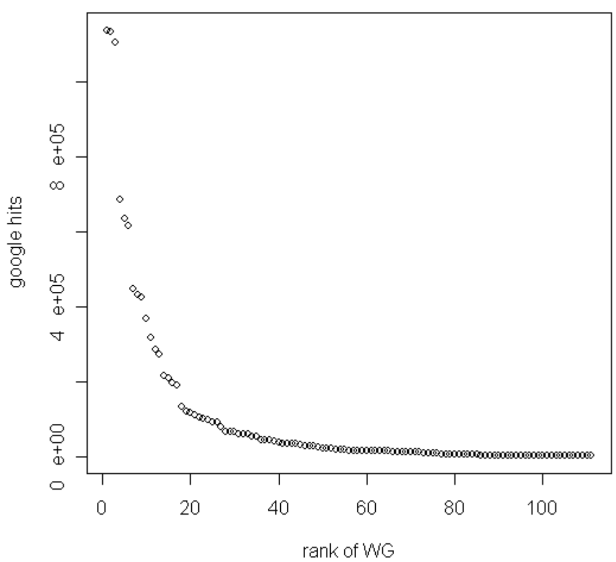

Figure 3. Ranking Working Groups by Impact

Specifically, according to Table 2, 80% of total Google hits come from only 16.2% of the WGs, which satisfies 80/20 rule of Pareto Principle[2]. Figure 4 shows the log-log rank distribution of Google hit of WG. It shows that a power law relationship may exist. Since the linear estimation doesn't correspond exactly, we separated the data set into two parts at the point when $\log X = 4$ (X equals 55^{th}) approximately. Figure 4 shows the result. The two straight lines represent least square estimation for the upper ranked data and lower ranked data with slope equals –1.38 and –6.26 respectively. This time, both parts fit much better.

% of Google hits	% of WGs
20	0.9
50	5.4
80	16.2

Table 2: Percent of Google hits and the Corresponding Percent of Working Groups

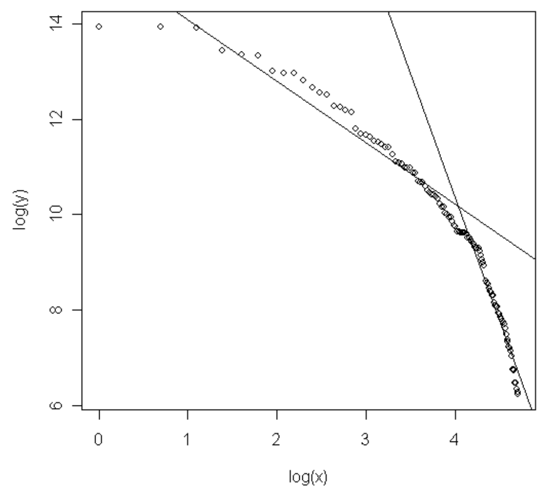

Figure 4. Log-Log Plot of Working Groups by Impact

V. RESULTS OF MULTIVARIATE ANALYSIS ON IMPACT

The analysis of our second data set focused on comparing the background information and author information against the quantifiable measures of standards. The analysis sought to identify influential variables that affect the success of a standard.

A second goal is determining variables that affect the impact of open standards. The variables include the length of the standards process, the parties involved, and the technical complexity of the standard. The impact of a standard is quantified through its popularity and various forms of citation indexes. Our

[2] The Pareto principle (also known as the 80-20 rule) states that for many phenomena, 80% of the consequences stem from 20% of the causes.

preliminary results showed that the length of the standard and the number of references were significant variables. Interestingly, the duration of the development process was not a significant factor. The regression results are in Table 3. A statistical summary and other related statistical results can be found in the Appendix.

Independent Variables	**DEPENDENT VARIABLES**	
	Log(Hits)	Log(S_Hits)
Log(Words)	1.47131 (0.37480) 0.000513 ***	0.9594 (0.2992) 0.00335 **
References	-0.11844 (0.04374) 0.011416 *	-0.1148 (0.0332) 0.00176 **
RFCs published before	0.03392 (0.01728) 0.059694 .	
Number of Authors		0.2316 (0.1050) 0.03578 *
Observation	32	
Adjusted R2	0.4033	0.399
p-value for F	0.0005331	0.0005873

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Table 3. Linear Regression Results:

Although our sample size is small, the two consistently significant variables of words and references are interesting. We consider the implications of these findings in the next section.

VI. IMPLICATIONS

The descriptive quantitative analysis found a power law distribution. The most prominent character of a power law is: a very small proportion of events has a very large occurrence, whereas a very large proportion of events is extremely rare. These types of distributions have been found in online contexts for the growth of the web [9]. In all these cases growth is modeled or considered analogous to growth in a biological community. While descriptive, more recent power law analysis has begun to explain why power laws might arise. For example, power law relationships have been found in networked communities, such as public companies, universities, and newspapers [15]. The work of scholars, such as Albert-László Barabási, Duncan Watts, and Bernardo Huberman have suggested that power law relationships occur in social systems where people can choose between many options. The result is an inequality that emerges. This inequality is not the choice of the IETF, but a natural property of providing so many different options. The implication here is that just a few standards will generate most of the impact in a "winner take all" scenario. There will be an inequality among standards. This suggests two implications for standards organizations.

First, standards organizations should happily accept that many standards will not become widely popular. This is just a simple property when you create lots of standards. This also carries implications. Open standards organizations should be flexible and adaptable in their approach towards the development of open standards. There will be some open standards that will require special guidance or rules because of their enormous impact, and vice-versa. After all, standards that are likely to have a high impact are often recognizable during the development process. They usually have more participants, are longer, and have more divisive debates. These results agree with our regressions that found longer standards generated higher impact. Standards organizations should not be afraid to play politics by instituting different rules or procedures to address problems during the standards development.

Secondly, standards organizations should recognize that many high quality and potentially useful standards may be overlooked. Nevertheless, standards organizations may wish to push against this and try to flatten the curve by encouraging the wider adoption of more standards. For example, the top 5% standards are well known and need little promotion, but standards groups could highlight the importance of the next 10% in an effort to increase their visibility and hopefully their adoption.

The results also showed that the length of a standards (akin to technical complexity of a standard) and the number of references were significant variables for the impact of a standards across multiple measures. The length of an IETF standard determined by the number of words is related to several factors. As a starting point, the IETF's norms and process suggest that standards be as compact and narrow as possible. Consequently, only standards that cover a large amount of technical terrain are allowed to be long. In Simcoe's analysis, he used the length of the standard as a proxy for technological complexity (2004). He argued that longer standards arise to solve complex problems and multiple points of view. Consequently, standards that are longer tend to have more participants, more discussion of the issues, more divisive debates, and technically more complex subject matter. This greater attention during the development process then translates into a greater impact for the standard.

The other variable that was consistently significant across multiple measures of impact was the number of references. The value for references was negative, suggesting the more references a standards has, the less impact it will have. An explanation for this result is that standards that are incremental in nature are likely to have a large of set of references. In contrast, innovative and novel work are less likely to have a large body of work to reference.

Interestingly, the duration of the development process was not significant. Earlier empirical research by Simcoe found that the IETF standards process has lengthened over time [21]. This work has proved instrumental for those seeking to reform the IETF standards process by reducing the duration of the development process. However, our results encourage such reforms, because there appears to be no relationship between a longer development process and the impact for a standard.

VII. CONCLUSION

The goal was to conduct empirical research on open standards. After all, open standards are widely praised, but there is little scholarly research on this topic. We found inequalities in the impact of open standards that suggest a power law relationship. The implications are that standard organizations need to recognize this property and shift their strategies during the development process. This may include treating potential high impact standards with different rules as well as focusing promotional efforts on "second-tier" standards.

The analysis of the impact of open standards found that the duration of the development process does not affect the impact of a standard. This finding carries significant policy implications as reforms are currently underway to speed up the IETF development process. Our results support these reforms as they should not influence the impact of the standards. Our analysis did find the length of the standard document (number of words) as a crucial factor affecting the impact of the standard. The length of a standard often reflects multiple participants, divisiveness, and technical complexity, all of which suggest greater interest during the development and consequently a higher impact.

The results here have led us to expand our data set. The original data set focused on 32 randomly selected standards in a set time period. Based on these results, we are going to expand the data set to all the IETF standards. This should allow us to broaden and deepen insights into open standards and the key variables that influence the impact of open standards.

Our future work may consider two other consortia, the World Wide Web Consortium (W3C) and Organization for the Advancement of Structured Information Standards (OASIS). These are all well regarded consortia and have different spheres of expertise. They also each were developed at different periods in time. The W3C was inspired by the IETF, but setup differently to avoid some of the perceived problems within the W3C. Similarly, OASIS was inspired by the W3C and IETF, but with different procedures to avoid problems that had occurred within the IETF and the W3C.

ACKNOWLEDGEMENTS

This material is based on work supported by the National Science Foundation under Grant Nos. 0429217. The authors would like to thank Jiang Li and Andrew Kennis for their research assistance.

REFERENCES

Besen, S. M., & Farrell, J. (1994). Choosing How to Compete: Strategies and Tactics in Standardization. *Journal of Economic Perspectives, 8*(2), 117-131.

Bradner, S. (1998, September). RFC 2418: IETF Working Group Guidelines and Procedures. Retrieved December 12, 2006, from http://www.faqs.org/rfcs/rfc2418.html

Committee for Economic Development. (2006). *Open Standards, Open Source, and Open Innovation: Harnessing the Benefits of Openness*. Washington, DC: Committee for Economic Development.

Crocker, D. (1993). Making Standards the IETF Way. *Standardview, 1*(1), 48-56.

David, P. A. (1985). Clio and the Economics of QWERTY. *American Economic Review, 75*(5), 332-337.

David, P. A., & Greenstein, S. (1990). The Economics of Compatibility Standards: An Introduction to Recent Research. *Economics of Innovations and New Technology, 1*(1), 3-41.

David, P. A., & Steinmueller, W. E. (1994). Economics of Compatibility Standards and Competition in Telecommunication Networks. *Information Economics and Policy, 6*, 217-241.

Farrell, J., & Saloner, G. (1988). Coordination Through Committees and Markets. *The RAND Journal of Economics, 19*(2), 235-252.

Huberman, B. A., & Adamic, L. A. (1999). Growth dynamics of the World-Wide Web. *Nature, 399*, 130.

Krechmer, K. (2006). The Meaning of Open Standards. *The International Journal of IT Standards and Standardization Research, 4*(1), 43-61.

Lehr, W. (1995). Compatibility Standards and Interoperability: Lessons from the Internet. In B. Kahin & J. Abbate (Eds.), *Standards Policy for Information Infrastructure* (pp. 121-147). Cambridge, MA: MIT Press.

Lemley, M. A. (2002). Intellectual Property Rights and Standard-Setting Organizations. *California Law Review, 90*, 1889-1980.

Lohr, S. (2005, September 9). Plan by 13 Nations Urges Open Technology Standards. *New York Times*.

Office of Government Commerce. (2004). *Open Source Software: Use within UK Government*.

Pennock, D. M., Flake, G. W., Lawrence, S., Glover, E. J., & Giles, C. L. (2002). Winners Don't Take All: Characterizing the Competition for Links on the Web. *Proceedings of the National Academy of Sciences, 99*, 5207-5211.

Shah, R. C., & Kesan, J. P. (2003). Manipulating the Governance Characteristics of Code. *Info, 5*(4), 3-9.

Shapiro, C., & Varian, H. R. (1999a). *Information Rules: A Strategic Guide to the Network Economy*. Boston, MA: Harvard Business School Press.

Shapiro, C., & Varian, H. R. (1999b). The Art of Standards Wars. *California Management Review, 41*(2), 8-32.

Simcoe, T. (2003). Committees and the Creation of Technical Standards.

Simcoe, T. (2004). Delays and de Jure Standards: What Caused the Slowdown in Internet Standards Development.

Tassey, G. (2005). Underinvestment in Public Good Technologies. *Journal of Technology Transfer, 30*, 90-94.

Updegrove, A. (1995). Standard Setting and Consortium Structures. *StandardView, 3*(4), 143-147.

West, J. (2006). The Economic Realities of Open Standards: Black, White and Many Shades of Gray. In S. Greenstein & V. Stango (Eds.), *Standards and Public Policy*. Cambridge: Cambridge University Press.

Zilles, S. (2004). Open Specifications: Mostly Open Standards. In S. Bolin (Ed.), *The Standards Edge: Dynamic Tension* (pp. 265-270). Ann Arbor, MI: Bolin Communications.

APPENDIX

RFC	Title
3031	Multiprotocol Label Switching Architecture
3074	DHC Load Balancing Algorithm
3119	A More Loss-Tolerant RTP Payload Format for MP3 Audio
3144	Remote Monitoring MIB Extensions for Interface Parameters Monitoring
3161	Internet X.509 Public Key Infrastructure Time-Stamp Protocol (TSP)
3195	Reliable Delivery for syslog
3225	Indicating Resolver Support of DNSSEC
3246	An Expedited Forwarding PHB (Per-Hop Behavior)
3265	Session Initiation Protocol (SIP)-Specific Event Notification
3268	Advanced Encryption Standard (AES) Ciphersuites for Transport Layer Security (TLS)
3320	Signaling Compression (SigComp)
3339	Date and Time on the Internet: Timestamps
3428	Session Initiation Protocol (SIP) Extension for Instant Messaging
3436	Transport Layer Security over Stream Control Transmission Protocol
3440	Definitions of Extension Managed Objects for Asymmetric Digital Subscriber Lines
3503	Message Disposition Notification (MDN) profile for Internet Message Access Protocol (IMAP)
3516	IMAP4 Binary Content Extension
3530	Network File System (NFS) version 4 Protocol
3554	On the Use of Stream Control Transmission Protocol (SCTP) with IPsec
3578	Mapping of Integrated Services Digital Network (ISDN) User Part (ISUP) Overlap Signaling to the Session Initiation Protocol (SIP)
3628	Policy Requirements for Time-Stamping Authorities (TSAs)
3641	Generic String Encoding Rules (GSER) for ASN.1 Types
3646	DNS Configuration options for Dynamic Host Configuration Protocol for IPv6 (DHCPv6)
3664	The AES-XCBC-PRF-128 Algorithm for the Internet Key Exchange Protocol (IKE)
3728	Definitions of Managed Objects for Very High Speed Digital Subscriber Lines (VDSL)
3767	Securely Available Credentials Protocol
3810	Multicast Listener Discovery Version 2 (MLDv2) for IPv6
3821	Fibre Channel Over TCP/IP (FCIP)
3853	S/MIME Advanced Encryption Standard (AES): Requirement for the Session Initiation Protocol (SIP)
3860	Common Profile for Instant Messaging (CPIM)
3863	Presence Information Data Format (PIDF)
3909	Lightweight Directory Access Protocol (LDAP) Cancel Operation

Table 1. Selected IETF Open Standards

Variables	**Minimum**	**Maximum**	**Mean**	**Std.Dev.**	
				Before transformation	**After transformation**
Number of authors	1	9	2.66	2.09	
Commercial authors	1	9	2.56	2.05	
RFCs published before	0	65	12.84	15.31	
Total RFCs published	1	98	22.44	23.13	
Duration of development process	122	1667	686.25	363.71	0.26
Number of Drafts	1	16	7.22	3.66	
Words	841	79716	9322.94	14339.08	0.44
Pages	3	274	32.34	48.88	
References	1	40	11.69	8.65	
References citing other RFCs	1	29	8.78	6.93	
Words in References	22	977	236.19	199.47	0.34
Hits	206	69100	9555.25	16050.21	0.81
Scholar Hits	4	1240	83.66	221.39	0.62
Citations from other RFCs	0	36	5.31	7.86	
References in patents	0	18	0.6	3.18	
IPRdisclosure	0	2	0.19	0.47	

Table 2. Statistical Summary of the Data

	Log (Words)	References	RFCbefore	Author	Drafts
Log (Words)	1.00	0.72	0.17	0.43	0.42
References	---	1.00	0.02	0.27	0.31
RFCbefore	---	---	1.00	---	---
Author	---	---	---	1.00	---
Drafts	---	---	---	---	1.00

Table 3. Correlation Matrix of Independent Variables

	Hits	S_Hits	otherRFC	ref_in_patent	IPRdisclosure
Hits	1.00	0.83	0.88	0.68	0.46
S_Hits	---	1.00	0.86	0.96	0.11
OtherRFC	---	---	1.00	0.73	0.26
ref_in_patent	---	---	---	1.00	-0.08
IPRdisclosure	---	---	---	---	1.00

Table 4 Correlation between the dependent variables:

	Hits	S_Hits	otherRFC	ref_in_patent	IPRdisclosure
References	-0.09	-0.11	-0.04	-0.16	-0.08

Table 5. Correlation between References and Dependent Variables: